

UNIVERSIDAD DE CANTABRIA

Escuela de Doctorado de la Universidad de Cantabria

Programa de Doctorado en Ingeniería Civil



Tesis doctoral

Ph.D thesis

**MEJORA EN LA RECOLECCIÓN DE DATOS Y LA MODELIZACIÓN DE LA
CALIDAD PERCIBIDA EN SISTEMAS DE TRANSPORTE PÚBLICO**

**IMPROVING DATA COLLECTION AND MODELLING OF PERCEIVED
QUALITY IN PUBLIC TRANSPORT SYSTEMS**

Autor

Author

ENEKO ECHANIZ BENEITEZ

Directores

Supervisors

LUIGI DELL'OLIO

ÁNGEL IBEAS PORTILLA

Santander, 2020

In memoriam

RESUMEN

Estudiar la calidad percibida en el transporte público es una forma de definir políticas efectivas enfocadas a mejorar el mismo. La satisfacción general de los clientes con respecto a un sistema de transporte público depende principalmente de dos factores: cuán satisfechos están estos con los diferentes aspectos que componen el servicio y cuán importante es cada uno de esos aspectos. La forma habitual de obtener esta información es mediante encuestas de satisfacción. Los datos obtenidos mediante las encuestas permiten estimar modelos que definen aquellos aspectos del servicio que más influyen en la percepción de la calidad de los usuarios. De esta forma, es posible establecer prioridades de inversión para los operadores o autoridades de los servicios.

La recopilación de datos para obtener información sobre la satisfacción de los clientes con los servicios de transporte público es muy lenta y costosa. Las encuestas destinadas a obtener la satisfacción de los clientes son considerablemente largas, lo que hace, que muchas encuestas no puedan completarse dentro del tiempo de viaje a bordo de los pasajeros/encuestados. El bajo rendimiento en el proceso de realización de encuestas requiere de un diseño de encuesta más eficiente que facilite la obtención de los datos.

Otro aspecto importante es la capacidad de los modelos estimados para representar la realidad. La interpretación de un modelo constituye la parte esencial para poder proponer mejoras en un servicio. Cuanto mejor sea el modelo estimado, mejores serán las conclusiones derivadas del mismo. Por esta razón, la mejora de los modelos supone un aspecto a considerar si se quiere aumentar la eficacia de las actuaciones destinadas a mejorar el servicio.

La presente tesis se encuentra constituida por un compendio de artículos compuesto por un total de 4 artículos. Todos ellos se engloban dentro del estudio de la calidad percibida en el transporte público, enfocándose en mejorar las dos cuestiones mencionadas anteriormente. Por una parte, los artículos se centran en mejorar el rendimiento de las encuestas proponiendo mejoras en el método utilizado actualmente o proponiendo un método alternativo más eficiente. Por otra parte, se mejoran los modelos utilizados anteriormente y se facilita su interpretación.

El primer artículo propone una metodología para considerar la importancia que los usuarios atribuyen a los distintos atributos del servicio dentro de los modelos de elección discreta estimados. Para la obtención de datos, se propone una encuesta de atributos basada en un ranking en dos etapas. En primer lugar, se clasifican los atributos pertenecientes a un grupo dentro de cada grupo. En segundo lugar, se clasifica cada grupo según su importancia con respecto a los demás grupos. Se proponen una serie de modelos ordered probit sucesivos que consideran, además de la importancia de los atributos, las variaciones sistemáticas y aleatorias en los gustos de los usuarios. La importancia de los atributos se incluye en los modelos mediante una ponderación de las variables. En el artículo se concluye que al aumentar la complejidad de los modelos se mejora su capacidad de representar la realidad. Sin embargo, llega un momento en que el esfuerzo necesario para obtener datos suficientes para alimentar la complejidad de los modelos no es eficiente y el tiempo empleado no se compensa con las mejoras en las predicciones.

El segundo artículo desarrolla un estudio para saber si es posible reducir la cantidad de información reunida en cada encuesta sin comprometer los resultados. Para abordar esta cuestión de investigación, se realiza un análisis comparativo de distintos modelos Ordered Probit: uno basado en los datos completos y otro basado en un conjunto de datos incompleto. En el caso de este último, la información faltante se imputa utilizando tres métodos distintos: Utilización de modas, imputaciones únicas utilizando modelos predictivos e imputación múltiple. Los resultados de la estimación muestran que el modelo parcial que utiliza el método de imputación múltiple se comporta de manera similar al modelo que se basa en el estudio completo.

El tercer artículo presenta un método combinado cualitativo y cuantitativo enfocado a determinar el nivel de importancia que los usuarios dan a los diferentes atributos de un servicio ferroviario. La identificación de estas preferencias permitirá establecer políticas que puedan aumentar la calidad y la demanda de este modo de transporte. El análisis cualitativo se basa en un proceso de participación a través de grupos focales y entrevistas en profundidad. A partir de estas fuentes se ha obtenido un diagnóstico del sistema y de los principales atributos que pueden influir en la demanda. La parte cuantitativa de la metodología se basa en la realización de una encuesta para determinar la importancia de los distintos atributos basada en un método de encuesta Best-Worst (Caso 1). Este método ha permitido determinar que los atributos relacionados con el sistema tarifario, el tiempo de viaje y la intermodalidad son los más importantes para aumentar la calidad del transporte ferroviario. En cambio, los atributos menos importantes son los relacionados con servicios adicionales muy específicos. Al mismo tiempo, se ha observado que la importancia varía según la frecuencia de uso del ferrocarril.

El cuarto artículo tiene por objeto verificar la posibilidad de sustituir el método tradicional de análisis de calidad percibida por un método más eficaz mediante el uso de encuestas del tipo Best-Worst. Para esta investigación se ha realizado una segunda encuesta de satisfacción a los usuarios del transporte público de la ciudad de Santander (España). Los resultados muestran que el nivel de satisfacción obtenido con estos métodos alternativos es notablemente similar. La importancia relativa de cada uno de los atributos que ofrecen los dos métodos difiere, y el método Best-Worst muestra resultados más intuitivos y coherentes con la literatura existente en este ámbito. Adicionalmente, se ha desarrollado un modelo de regresión para obtener la satisfacción del cliente de cada atributo del servicio a partir de los resultados de la modelización de Best-Worst.

ABSTRACT

Studying perceived quality in a public transport system is a way of defining effective policies aimed at improving it. Customer overall satisfaction towards a public transport system depends mainly on two factors: how satisfied they are with different aspects that make up the service and how important each of the service aspects is to the customer. The common way to obtain this information is through satisfaction surveys. The data obtained through the surveys allow the researchers to estimate models that define those aspects of the service that most influence the perception of user satisfaction. Thus, it is possible to establish investment priorities for service operators or authorities.

Collecting data to obtain information on customer satisfaction in public transport services is very time consuming and expensive. Customer satisfaction surveys are quite long, which means that many surveys cannot be completed within the on board travel time of the passengers/respondents. Low performance in the survey process requires a more efficient survey design that facilitates data collection.

Another important aspect is the ability of the estimated models to represent reality. A good interpretation of a model is essential to propose accurate improvements on a service. The better the estimated model, the better the conclusions derived from it. For this reason, the improvement of the models is an aspect to consider in order to increase the effectiveness of the actions aimed to improve the service.

The present thesis consists of a compendium of 4 articles. All of them are included within the study of perceived quality in public transport, focusing on improving the two aspects mentioned above. On the one hand, the articles focus on improving the performance of the surveying process by proposing improvements to the currently used method, or by proposing a more efficient alternative method. On the other hand, the models used in previous studies are improved and their interpretation is facilitated.

The first article proposes a method to consider the importance that users place to different service attributes within the estimated discrete choice models. To obtain the data, a satisfaction survey was carried out among public transport users in the city of Santander (Spain). In the survey, a two-stage ranking was proposed. First, the attributes belonging to a group were classified based on their importance level. After finishing all groups, the groups itself were ranked according to their importance. A series of successive ordered probit models are estimated. Models consider systematic and random variations in user tastes, in addition to the importance of the attributes. The importance of the attributes is included in the models through a weighting of the variables. The article concludes that increasing the complexity of the models improves their ability to represent reality. However, there is a stage when the effort required to obtain sufficient data to feed the complexity of the models is not efficient and the time spent is not compensated by improvements in the predictions.

The second article develops a study to see if it is possible to reduce the amount of information collected in each survey without compromising the results. To address this research question, a comparative analysis of different Ordered Probit models is carried out: one based on the complete dataset and another based on an incomplete dataset. In the

case of the latter, the missing information was imputed using three different methods: using the *mode*, single imputations using predictive models and multiple imputation. The results of the estimation show that the partial model using the multiple imputation method behaves similarly to the model based on the full dataset.

The third article presents a combined qualitative and quantitative method aimed at determining the level of importance that users give to the different attributes of a railway service. The identification of these preferences make possible to define policies that can increase quality and demand for this transport mode. The qualitative analysis is based on a process of participation through focus groups and in-depth interviews. From these sources, a diagnosis of the system and the main attributes that can influence demand have been obtained. The quantitative part of the methodology is based on a survey to determine the importance of the different attributes based on a Best-Worst (case 1) survey method. This method has determined that the attributes related to the fare system, travel time and intermodality are the most important to increase the quality of rail transport. In contrast, the less important attributes were those related to very specific additional services. At the same time, it has been observed that the importance of the attributes varies according to the frequency of use of the railway.

The fourth article aims to verify the possibility of replacing the traditional method of satisfaction analysis by a more efficient method using Best-Worst type survey. For this research, a second satisfaction survey was carried out among public transport users in the city of Santander (Spain). The results show that the level of satisfaction obtained with these alternative methods is remarkably similar. However, the relative importance of each of the attributes obtained by the two methods is different, although a high correlation is observed. The Best-Worst method shows results that are more intuitive and consistent with the existing literature. Additionally, a regression model has been developed to obtain the satisfaction level of each service attribute based on the Best-Worst data.

AGRADECIMIENTOS

Me gustaría empezar agradeciendo a las entidades con cuya financiación ha sido posible realizar esta tesis. Por una parte, a los proyectos de investigación TRA2015-69903-R, financiado por el Ministerio de Economía e Industria del Gobierno de España y cofinanciado con fondos FEDER; SETA-project (grant agreement N° 688082), financiado por el Consejo Europeo de Investigación (ERC) bajo el programa de la Unión Europea Horizonte 2020; y NEAR2050 (Project Reference: 730838, Call Reference S2R-OC-CCA-01-2015), financiado por la iniciativa europea Shift2Rail. Por otra parte, al Ministerio de Educación, Cultura y Deporte del Gobierno de España por concederme la ayuda para la formación de profesorado universitario FPU15 / 02990 y por la ayuda a la movilidad para estancias breves y traslados temporales EST18/00547 que permitió financiar una de las estancias realizadas en esta tesis. Ambas ayudas pertenecen actualmente al Ministerio de Ciencia e Innovación.

A mis directores, los catedráticos Luigi dell'Olio y Ángel Ibeas, les agradezco la ayuda ofrecida para guiarme a lo largo de este inicio de mi carrera investigadora, especialmente en los primeros meses, cuando más difícil es tomar decisiones. A mis compañeros y excompañeros de trabajo del grupo de Investigación de Sistemas de Transporte (GIST) y posteriormente SUM+LAB (Movilidad sostenible e ingeniería ferroviaria): Andrés Rodríguez, Gonzalo Antolín, Javier Fernández, Juan Benavente, Mikel Cerrada, Rebeca Bolado, Roberto Sañudo, Rosa Barreda, Rubén Cordera, Sara Ezquerro y José Luís Moura, que en mayor o menor medida habéis contribuido a que esta tesis pudiese llevarse a cabo, gracias. A Borja Alonso, gracias por haberme ayudado en la labor docente de estos últimos años y a ser mejor profesor.

Gracias al profesor David Henser, director del Institute of Transport and Logistics Studies (ITLS) de la University of Sydney (Sídney, Australia), por permitirme realizar dos estancias con ellos. Igualmente, gracias a todos los trabajadores del ITLS por ayudarme en todo lo que necesitaba y hacer mi vida más fácil. En especial, muchas gracias a Chinh Ho por ayudarme enormemente en el desarrollo de mi investigación, por revisar mi trabajo y por los consejos que me permitieron publicar dos de los cuatro artículos incluidos en esta tesis. Sin su inestimable ayuda esta tesis no hubiese podido realizarse.

Aunque la investigación es importante, quiero dar las gracias a Sonali Tripathi y a Göran Smith por esos momentos fuera de la oficina, por los planes alternativos, gracias por hacer de mi tiempo en Sydney una experiencia inolvidable.

A Janet Porter, porque lo que empezó como una relación profesional se volvió una gran amistad. Gracias por ayudarme en todo lo que necesitaba. Gracias por las conversaciones, las barbacoas y las cervezas después del trabajo. A Nina y Maisie que las echo muchísimo de menos. Y no me olvido de Tyrone McNulty, gracias por acompañarnos en todo momento. Siempre nos quedará el Bar 44.

Estos cuatro años y medio en Santander no hubiesen sido lo mismo si no fuera por toda esa gente que ha estado ahí en una ciudad que considero ahora como mi segunda casa. Quiero agradecer a los miembros del equipo de natación del ACN Marisma por los buenos ratos que hemos pasado juntos en la piscina y fuera de ella. Especialmente, gracias a

Artur Martí, Diego Martínez, Iván Polo, Jaime Teja y Miguel Ángel Anievas, por enseñarme Cantabria y sus gentes.

A mis amigos de toda la vida, que siempre han estado ahí.

Lucía, gracias por estar ahí en los buenos momentos y en los malos, especialmente en los malos. Gracias por aguantarme todo este tiempo que no es siempre fácil. Te quiero.

A mi familia, a mis padres y hermanos.

Eskerrik asko.

ACKNOWLEDGEMENT

I would like to start by thanking the entities with whose financing this thesis has been possible. On the one hand, to the research projects: TRA2015-69903-R, financed by the Spanish Ministry of Economy and Industry and co-financed with ERDF funds; SETA-project (grant agreement N° 688082), financed by the European Research Council (ERC) under the European Union Horizon 2020 program; and NEAR2050 (Project Reference: 730838, Call Reference S2R-OC-CCA-01-2015), financed by the European initiative Shift2Rail. On the other hand, to the Ministry of Education, Culture and Sport of the Spanish Government for granting me the scholarship for developing my thesis FPU15 / 02990 and for the mobility funds EST18/00547 that allowed to cover all the expenses of one of the stays made in this thesis. Both grants currently belong to the Ministry of Science and Innovation.

I would like to thank my directors, Professors Luigi dell'Olio and Angel Ibeas, for their help in guiding me through this early stage of my research career, especially in the first months, when it is most difficult to make decisions. To my colleagues and former colleagues of the Transport Systems Research Group (GIST) and later SUM+LAB (Sustainable mobility and railways engineering): Andrés Rodríguez, Gonzalo Antolín, Javier Fernández, Juan Benavente, Mikel Cerrada, Rebeca Bolado, Roberto Sañudo, Rosa Barreda, Rubén Cordera, Sara Ezquerro and José Luís Moura, who to a greater or lesser extent have contributed to this thesis, thank you. To Borja Alonso, thank you for helping me these last years of teaching and for making me a better teacher.

Thanks to Professor David Henser, Director of the Institute of Transport and Logistics Studies (ITLS) at the University of Sydney (Sydney, Australia), for allowing me to do two stays with them. Also, thanks to all the ITLS workers for helping me in everything I needed and making my life easier. Special thanks to Chinh Ho for helping me enormously in the development of my research, for reviewing my work and for the advices that allowed me to publish two of the four articles included in this thesis. Without his invaluable help this thesis would not have been possible.

Although the research is important, I would like to thank Sonali Tripathi and Göran Smith for those moments away from the office, for the alternative plans, and for making my time in Sydney an unforgettable experience.

To Janet Porter, because what started as a professional relationship turned into a great friendship. Thank you for helping me with everything I needed. Thank you for the conversations, the barbecues and the beers after work. To Nina and Maisie, I miss you both so much. And I don't forget Tyrone McNulty, thanks for being there for us all the time. We'll always have Bar 44.

These four and a half years in Santander wouldn't have been the same if it wasn't for all those people who have been there in a city I now consider my second home. I would like to thank the members of the ACN Marisma swim team for the good times we have spent together in the pool and out of it. Especially, thanks to Artur Martí, Diego Martínez, Iván Polo, Jaime Teja and Miguel Angel Anievas, for showing me Cantabria and its people.

Lucia, thank you for being there in the good and the bad times, especially in the bad.
Thank you for taking care of me all this time which is not always easy. I love you.

To my lifelong friends, who have always been there.

To my family, to my parents and brothers.

Eskerrik asko.

ÍNDICE DE CONTENIDO

RESUMEN	I
ABSTRACT	III
AGRADECIMIENTOS	V
ACKNOWLEDGEMENT	VII
ÍNDICE DE CONTENIDO.....	IX
ÍNDICE DE FIGURAS	XI
ÍNDICE DE TABLAS	XIII
CAPÍTULO 1: INTRODUCCIÓN Y OBJETIVOS	1
1.1 NORMATIVA DE REFERENCIA	3
1.2 ANTECEDENTES	3
1.3 MOTIVACIÓN Y OBJETIVOS	5
1.4 RELACIÓN ENTRE ARTÍCULOS DEL COMPENDIO	6
1.5 ESTRUCTURA DE LA TESIS	8
CAPÍTULO 2: SÍNTESIS DEL ESTADO DEL ARTE	11
2.1 CALIDAD EN EL TRANSPORTE PÚBLICO	13
2.2 ENCUESTAS DE SATISFACCIÓN	14
2.3 ENFOQUE METODOLÓGICO DEL ANÁLISIS DE LA CALIDAD PERCIBIDA	17
CAPÍTULO 3: METODOLOGÍA EMPLEADA.....	21
3.1 MODELOS DE ELECCIÓN DISCRETA.....	23
3.1.1 <i>Modelo Logit Multnomial</i>	24
3.1.2 <i>Modelo Mixed Logit (Parámetros aleatorios)</i>	25
3.1.3 <i>Modelo MNL para encuestas Best-Worst</i>	26
3.2 MODELOS DE DATOS ORDENADOS.....	27
3.2.1 <i>Modelos de datos ordenados con parámetros aleatorios</i>	29
3.3 DISEÑO DE LA ENCUESTAS	30
3.3.1 <i>Encuestas de Satisfacción</i>	30
3.3.2 <i>Encuestas Best-Worst</i>	31
3.4 TAMAÑO MUESTRAL	34
3.5 MÉTODOS UTILIZADOS EN CADA ARTÍCULO	36
3.5.1 <i>Artículo 1</i>	36
3.5.2 <i>Artículo 2</i>	38
3.5.3 <i>Artículo 3</i>	39
3.5.4 <i>Artículo 4</i>	41
CAPÍTULO 4: COMPENDIO DE ARTÍCULOS	45
4.1 ARTÍCULO 1: MODELLING PERCEIVED QUALITY FOR URBAN PUBLIC TRANSPORT SYSTEMS USING WEIGHTED VARIABLES AND RANDOM PARAMETERS.....	47
4.1.1 <i>Indicadores bibliométricos de la revista de publicación del artículo 1 en el JCR</i>	47
4.1.2 <i>Transcripción del artículo 1: Modelling perceived quality for urban public transport systems using weighted variables and random parameters</i>	48
4.2 ARTÍCULO 2: MODELLING USER SATISFACTION IN PUBLIC TRANSPORT SYSTEMS CONSIDERING MISSING INFORMATION.....	71
4.2.1 <i>Indicadores bibliométricos de la revista de publicación del artículo 2 en el JCR</i>	71
4.2.2 <i>Transcripción del artículo 2: Modelling user satisfaction in public transport systems considering missing information</i>	72
4.3 ARTÍCULO 3: ADDRESSING THE IMPORTANCE OF SERVICE ATTRIBUTES IN RAILWAYS.....	89
4.3.1 <i>Indicadores bibliométricos de la revista de publicación del artículo 3 en el JCR</i>	89

4.3.2	<i>Transcripción del artículo 3: Addressing the importance of service attributes in railways</i>	90
4.4	ARTÍCULO 4: COMPARING BEST-WORST AND ORDERED LOGIT APPROACHES FOR USER SATISFACTION IN TRANSIT SERVICES	115
4.4.1	<i>Indicadores bibliométricos de la revista de publicación del artículo 4 en el JCR.....</i>	115
4.4.2	<i>Transcripción del artículo 4: Comparing best-worst and ordered logit approaches for user satisfaction in transit services.....</i>	116
CAPÍTULO 5: RESULTADOS Y DISCUSIÓN		144
5.1	ENCUESTA DE SATISFACCIÓN TUS 2015	146
5.1.1	<i>Resultados socioeconómicos de la muestra</i>	146
5.1.2	<i>Resultados de la valoración de los atributos</i>	148
5.1.3	<i>Modelización considerando variaciones sistemáticas y aleatorias en los gustos</i>	149
5.1.4	<i>Comparativa de los modelos</i>	160
5.1.5	<i>Análisis de los resultados considerando datos faltantes.....</i>	161
5.2	ENCUESTA DE SATISFACCIÓN TUS 2017	168
5.2.1	<i>Resultados socioeconómicos de la muestra</i>	168
5.2.2	<i>Resultados de la valoración de los atributos</i>	169
5.2.3	<i>Modelización mediante modelos de elección discreta</i>	170
5.2.4	<i>Análisis de Importancia-Valoración (IPA: Importance-Performance Analysis).....</i>	176
5.2.5	<i>Modelos de regresión entre los distintos métodos de modelización</i>	178
5.3	ENCUESTA SOBRE LA IMPORTANCIA EN EL FERROCARRIL	180
5.3.1	<i>Análisis cualitativo</i>	180
5.3.2	<i>Resultados socioeconómicos de la muestra</i>	182
5.3.3	<i>Resultados de la modelización sobre la importancia de los atributos en el ferrocarril ...</i>	183
5.3.4	<i>Variación de la importancia de acuerdo al tipo de usuario.</i>	185
CAPÍTULO 6: CONCLUSIONES FINALES Y LÍNEAS DE INVESTIGACIÓN FUTURAS		189
6.1	CONCLUSIONES ESPECÍFICAS.....	191
6.2	CONCLUSIONES GENERALES	194
6.3	LÍNEAS DE INVESTIGACIÓN FUTURA	195
CAPÍTULO 7: CONCLUSIONS AND FUTURE LINES OF RESEARCH		199
7.1	SPECIFIC CONCLUSIONS	201
7.2	GENERAL CONCLUSIONS	204
7.3	FUTURE LINES OF RESEARCH	205
REFERENCIAS		209

ÍNDICE DE FIGURAS

FIGURA 1 CICLO DE LA CALIDAD DE SERVICIO. FUENTE: AENOR - UNE-EN 13816.....	13
FIGURA 2 EJEMPLO DE PREGUNTA BEST-WORST DE LA ENCUESTA DE IMPORTANCIA FERROVIARIA (ARTÍCULO 3)	41
FIGURA 3 EJERCICIO DE RESPUESTA DE LA ENCUESTA DE SATISFACCIÓN DE 2017 (ARTÍCULO 4).....	42
FIGURA 4 ARTÍCULO 2: FIGURE 1 – USERS’ SATISFACTION	83
FIGURA 5 ARTÍCULO 2: FIGURE 2 - NORMALIZED MODEL PARAMETERS: A COMPARISON OF MODELS WITH FULL VS. PARTIAL DATA	86
FIGURA 6 ARTÍCULO 3: FIGURE 1. METHODOLOGY PROPOSED FOR STUDYING THE NEEDS OF TRANSPORT USERS.....	93
FIGURA 7 ARTÍCULO 3: FIGURE 2 SHOWS AN EXAMPLE OF A BW CHOICE SCENARIO BETWEEN FOUR ATTRIBUTES.....	102
FIGURA 8 ARTÍCULO 3: FIGURE 3. IMPORTANCE OF THE ATTRIBUTES ACCORDING TO THEIR NORMALIZED COEFFICIENT	106
FIGURA 9 ARTÍCULO 3: FIGURE 4. CORRELATION BETWEEN THE PARAMETERS OBTAINED IN THE MNL MODEL AND THE BEST–WORST OPERATION.....	107
FIGURA 10 ARTÍCULO 3: FIGURE 5. AVERAGE BEST–WORST VALUES AND MAXIMUM AND MINIMUM GIVEN BY THE RESPONDENTS BY TYPE OF ATTRIBUTE.....	108
FIGURA 11 ARTÍCULO 3: FIGURE 6. COMPARISON OF THE DIFFERENCE IN IMPORTANCE GIVEN TO THE ATTRIBUTES BY THE COMPLETE SAMPLE AND FREQUENT USERS	110
FIGURA 12 ARTÍCULO 3: FIGURE 7. COMPARISON OF THE DIFFERENCE IN IMPORTANCE GIVEN TO THE ATTRIBUTES BY THE COMPLETE SAMPLE AND WEEKLY USERS	111
FIGURA 13 ARTÍCULO 3: FIGURE 8. COMPARISON OF THE DIFFERENCE IN IMPORTANCE GIVEN TO THE ATTRIBUTES BY THE COMPLETE SAMPLE AND OCCASIONAL USERS/NONUSERS	112
FIGURA 14 ARTÍCULO 4: FIGURE 1 – EXAMPLE DATA COLLECTED FROM ONE CHOICE TASK BASED ON RATING VS. BW SCALES	121
FIGURA 15 ARTÍCULO 4: FIGURE 2 - COMPARISON BETWEEN BW (SATISFACTION) MODEL AND SATISFACTION RATING	130
FIGURA 16 ARTÍCULO 4: FIGURE 3 – COMPARISON BETWEEN BW (IMPORTANCE) AND ORDERED LOGIT MODELS	132
FIGURA 17 ARTÍCULO 4: FIGURE 4 – IMPORTANCE – PERFORMANCE ANALYSIS.....	134
FIGURA 18 COMPARATIVA DE LA VALORACIÓN DE LOS ATRIBUTOS CONSIDERANDO LA BASE DE DATOS COMPLETA Y CON DATOS FALTANTES.....	162
FIGURA 19 COMPARATIVA DE LOS PARÁMETROS ESTIMADOS EN LOS MODELOS CON DATOS FALTANTES.	166
FIGURA 20 COMPARACIÓN ENTRE LA PUNTUACIÓN DE SATISFACCIÓN Y LOS PARÁMETROS DEL MODELO MNL BASADO EN DATOS BW DE SATISFACCIÓN	174
FIGURA 21 COMPARACIÓN DE LOS PARÁMETROS DEL MODELO OL Y EL MNL BASADO EN DATOS BW DE IMPORTANCIA.....	175
FIGURA 22 ANÁLISIS IPA DE LA ENCUESTA DE SATISFACCIÓN DE 2017.....	177
FIGURA 23 COMPARATIVA ENTRE LA IMPORTANCIA DE LOS ATRIBUTOS FERROVIARIOS DE LA MUESTRA COMPLETA Y LOS USUARIOS FRECUENTES	186
FIGURA 24 COMPARATIVA ENTRE LA IMPORTANCIA DE LOS ATRIBUTOS FERROVIARIOS DE LA MUESTRA COMPLETA Y LOS USUARIOS SEMANALES.....	187
FIGURA 25 COMPARATIVA ENTRE LA IMPORTANCIA DE LOS ATRIBUTOS FERROVIARIOS DE LA MUESTRA COMPLETA Y LOS USUARIOS OCASIONALES O NO USUARIOS	188

ÍNDICE DE TABLAS

TABLA 1 EJEMPLO DE PREGUNTA BW CASO 1	32
TABLA 2 EJEMPLO DE PREGUNTA BW CASO 2.....	33
TABLA 3 EJEMPLO DE PREGUNTA BW CASO 3.....	33
TABLA 4 VARIABLES SOCIOECONÓMICAS OBTENIDAS EN LA ENCUESTA DE SATISFACCIÓN DE 2015 (ARTÍCULO 1)	37
TABLA 5 ATRIBUTOS DEL SERVICIO EVALUADOS EN LA ENCUESTA DE SATISFACCIÓN DE 2015 (ARTÍCULO 1)	37
TABLA 6 ATRIBUTOS INCLUIDOS EN LA ENCUESTA DE IMPORTANCIA FERROVIARIA (ARTÍCULO 3)	39
TABLA 7 ATRIBUTOS DE SERVICIO INCLUIDOS EN LA ENCUESTA DE SATISFACCIÓN DE 2017 (ARTÍCULO 4).....	41
TABLA 8 FACTORES DE IMPACTO DE LA REVISTA DEL ARTÍCULO 1	47
TABLA 9 RANKING DE LA REVISTA DEL ARTÍCULO 1	47
TABLA 10 NÚMERO DE CITAS DEL ARTÍCULO 1	47
TABLA 11 ARTÍCULO 1: TABLE 1 – PUBLIC TRANSPORT SERVICE ATTRIBUTES.....	50
TABLA 12 ARTÍCULO 1: TABLE 2 – CHARACTERISATION VARIABLES.....	51
TABLA 13 ARTÍCULO 1: TABLE 3– NUMBER OF SURVEYS.....	56
TABLA 14 ARTÍCULO 1: TABLE 4– BASE MODEL (BM)	57
TABLA 15 ARTÍCULO 1: TABLE 5– SYSTEMATIC VARIATIONS MODEL (SVM)	59
TABLA 16 ARTÍCULO 1: TABLE 6 – SYSTEMATIC VARIATIONS MODEL WITH WEIGHTED VARIABLES (SVMW)	61
TABLA 17 ARTÍCULO 1: TABLE 7– BASE MODEL WITH RANDOM PARAMETERS (RBM).....	62
TABLA 18 ARTÍCULO 1: TABLE 8 – SYSTEMATIC AND RANDOM VARIATIONS MODEL (SRVM).....	64
TABLA 19 ARTÍCULO 1: TABLE 9 – SYSTEMATIC AND RANDOM VARIATIONS MODEL WITH WEIGHTED VARIABLES (SRVMW)	66
TABLA 20 ARTÍCULO 1: TABLE 10 –COMPARISON OF MODELS.....	68
TABLA 21 FACTORES DE IMPACTO DE LA REVISTA DEL ARTÍCULO 2	71
TABLA 22 RANKING DE LA REVISTA DEL ARTÍCULO 2	71
TABLA 23 NÚMERO DE CITAS DEL ARTÍCULO 2	71
TABLA 24 ARTÍCULO 2: TABLE 1 - VARIABLES INCLUDED IN THE SURVEY	80
TABLA 25 ARTÍCULO 2: TABLE 2 - DESCRIPTIVE ANALYSIS	81
TABLA 26 ARTÍCULO 2: TABLE 3 - ORDERED PROBIT MODELS: PARAMETER ESTIMATES AND T-VALUES.....	84
TABLA 27 ARTÍCULO 2: TABLE 4 - VUONG TEST FOR NON-NESTED MODELS	87
TABLA 28 FACTORES DE IMPACTO DE LA REVISTA DEL ARTÍCULO 3	89
TABLA 29 RANKING DE LA REVISTA DEL ARTÍCULO 3	89
TABLA 30 NÚMERO DE CITAS DEL ARTÍCULO 3	89
TABLA 31 ARTÍCULO 3: TABLE 1. SWOT (STRENGTHS, WEAKNESSES, OPPORTUNITIES, AND THREATS) ANALYSIS: STRENGTHS AND WEAKNESSES EXTRACTED FROM FOCUS GROUPS (FGs) AND IN-DEPTH INTERVIEWS	97
TABLA 32 ARTÍCULO 3: TABLE 2. SWOT ANALYSIS: OPPORTUNITIES AND THREATS EXTRACTED FROM FGs AND IN-DEPTH INTERVIEWS	98
TABLA 33 ARTÍCULO 3: TABLE 3. ATTRIBUTES OBTAINED FROM FGs AND IN-DEPTH INTERVIEWS.....	99
TABLA 34 ARTÍCULO 3: TABLE 4. EXPERIMENTAL DESIGN OF THE DEFINITIVE STATED PREFERENCE (SP) SURVEY (THE CODES OF THE ATTRIBUTES CAN BE CONSULTED IN TABLE 3).....	101
TABLA 35 ARTÍCULO 3: TABLE 5. SOCIODEMOGRAPHIC CHARACTERISTICS OF RESPONDENTS	103
TABLA 36 ARTÍCULO 3: TABLE 6. MULTINOMIAL LOGIT (MNL) MODEL DERIVED FROM THE BEST–WORST EXERCISE	104
TABLA 37 ARTÍCULO 3: TABLE 7. MNL MODEL WITH INTERACTIONS BY RAILWAY USAGE LEVEL.....	109
TABLA 38 FACTORES DE IMPACTO DE LA REVISTA DEL ARTÍCULO 4	115
TABLA 39 RANKING DE LA REVISTA DEL ARTÍCULO 4	115
TABLA 40 NÚMERO DE CITAS DEL ARTÍCULO 4	115
TABLA 41 ARTÍCULO 4: TABLE 1 - RESPONDENTS' SOCIO-ECONOMIC INFORMATION.....	122
TABLA 42 ARTÍCULO 4: TABLE 2 – SATISFACTION RATINGS	123
TABLA 43 ARTÍCULO 4: TABLE 3 – ESTIMATION RESULTS OF BW AND ORDERED LOGIT MODELS (T-VALUE IN PARENTHESES).....	127

TABLA 44 ARTÍCULO 4: TABLE 4 – CORRELATION COEFFICIENTS.....	129
TABLA 45 ARTÍCULO 4: TABLE 5 – SATISFACTION REGRESSION MODEL.....	135
TABLA 46 ARTÍCULO 4: TABLE 6 – IMPORTANCE REGRESSION MODEL.....	136
TABLA 47 ARTÍCULO 4: TABLE 21- RANDOM PARAMETER ORDERED LOGIT	139
TABLA 48 ARTÍCULO 4: TABLE 22 MIXED LOGIT MODEL FOR SATISFACTION BW DATA	140
TABLA 49 ARTÍCULO 4: TABLE 23 MIXED LOGIT MODEL FOR IMPORTANCE BW DATA.....	142
TABLA 50 ARTÍCULO 4: TABLE 24 CORRELATION COEFFICIENTS OF MEANS OF RP MODELS.....	143
TABLA 51 ARTÍCULO 4: TABLE 25 CORRELATION COEFFICIENTS OF SCALE PARAMETERS OF RP MODELS	143
TABLA 52 RESULTADOS SOCIOECONÓMICOS DE LA ENCUESTA DE SATISFACCIÓN DE 2015	146
TABLA 53 RESULTADOS DE LA VALORACIÓN DE LA ENCUESTA DE SATISFACCIÓN DE 2015	148
TABLA 54 RESULTADOS DE ESTIMACIÓN DEL MODELO BASE (BM)	150
TABLA 55 RESULTADOS DE ESTIMACIÓN DEL MODELO CON INTERACCIONES (SVM)	151
TABLA 56 RESULTADOS DE ESTIMACIÓN DEL MODELO CON INTERACCIONES Y VARIABLES PONDERADAS (SVMW)	153
TABLA 57 RESULTADOS DE ESTIMACIÓN DEL MODELO BASE CON PARÁMETROS ALEATORIOS (RBM)	155
TABLA 58 RESULTADO DEL MODELO QUE CON VARIACIONES SISTEMÁTICAS Y ALEATORIAS EN LOS GUSTOS (SRVM)	157
TABLA 59 RESULTADO DEL MODELO CON INTERACCIONES, VARIABLES PONDERADAS Y PARÁMETROS ALEATORIOS (SRVMW)	159
TABLA 60 COMPARATIVA DE LOS MODELOS.....	161
TABLA 61 RESULTADOS DE LOS MODELOS CON DATOS FALTANTES ESTIMADOS	163
TABLA 62 TEST VUONG PARA MODELOS NO ANIDADOS	167
TABLA 63 RESULTADOS SOCIOECONÓMICOS DE LA MUESTRA	168
TABLA 64 VALORACIONES DE LA SATISFACCIÓN.....	169
TABLA 65 RESULTADOS DE LOS MODELOS OL Y MNL BASADOS EN BW	171
TABLA 66 CORRELACIÓN ENTRE LOS PARÁMETROS DE LOS MODELOS OL Y MNL BASADOS EN BW	173
TABLA 67 MODELO DE REGRESIÓN PARA EL NIVEL DE SATISFACCIÓN DE LOS ATRIBUTOS	179
TABLA 68 MODELO DE REGRESIÓN PARA LA IMPORTANCIA DE LOS ATRIBUTOS.....	179
TABLA 69 ANÁLISIS DAFO (DEBILIDADES, AMENAZAS, FORTALEZAS Y OPORTUNIDADES)	180
TABLA 70 RESULTADOS SOCIOECONÓMICOS DE LA ENCUESTA DE IMPORTANCIA FERROVIARIA	182
TABLA 71 MODELO MNL BASADO EN LOS DATOS BEST-WORST	184
TABLA 72 MODELO MNL CON INTERACCIONES POR FRECUENCIA DE USO DEL FERROCARRIL	185

“Empezar los capítulos con una cita es bastante molón. Tiene un punto cool. Un punto aspiracional. Hasta un punto de tío leído.”

“Starting the chapters with a quote is quite amazing. Has a cool point. An aspirational point. Even a point of a cultured guy.”

(Sanjuan, 2014)

CAPÍTULO 1: INTRODUCCIÓN Y OBJETIVOS

1.1 NORMATIVA DE REFERENCIA

La presente tesis doctoral se articula como una tesis por compendio de artículos previamente publicados, con mención internacional y se rige de acuerdo a las disposiciones generales definidas para realizar los estudios de Doctorado y la obtención del Título de Doctor de la Universidad de Cantabria. A su vez, la presente tesis se encuentra enmarcada dentro del Programa de Doctorado en Ingeniería Civil (PDIC), debiendo cumplir de igual manera con las disposiciones específicas definidas en dicho programa. A continuación, se enumera la normativa vigente:

- La Ley Orgánica 6/2001, de 21 de diciembre, de Universidades y sus normas de desarrollo, modificada por la Ley Orgánica 4/2007, de 12 de abril.
- El Real Decreto 99/2011, de 28 de enero, por el que se regulan las enseñanzas oficiales de Doctorado.
- El Real Decreto 195/2016, de 13 de mayo, por el que se establecen los requisitos para la expedición del Suplemento Europeo al Título Universitario de Doctor y se modifica el Real Decreto 99/2011.
- Los Estatutos de la UC, aprobados por Decreto 26/2012, de 10 de mayo (Boletín Oficial de Cantabria de 17 de mayo de 2012).
- Otras disposiciones sobre doctorado promulgadas por el Estado, adaptadas a la Universidad de Cantabria.
- Normativa de gestión académica de los estudios de doctorado regulados por el Real Decreto 99/2011, aprobada por el Consejo de Gobierno de la Universidad de Cantabria el 4 de marzo de 2015.
- Normativa para la elaboración de tesis doctorales como compendio de artículos dentro del Programa de Doctorado en Ingeniería Civil (PDIC) de la Escuela de Doctorado de la Universidad de Cantabria (EDUC). Aprobada por la Comisión académica del Programa de Doctorado en Ingeniería civil el 20 de mayo de 2016 y ratificada por el Comité de Dirección de la Escuela de Doctorado de la Universidad de Cantabria el 14 de junio de 2016.

1.2 ANTECEDENTES

La presente tesis se encuentra enmarcada dentro de la línea de investigación denominada “Estudio de la calidad en el transporte público” desarrollada inicialmente en el Grupo de Investigación de Ingeniería de Sistemas de Transporte (GIST) y posteriormente en el grupo de investigación en Movilidad Sostenible e Ingeniería Ferroviaria (SUM⁺LAB). Esta línea de investigación ha seguido desarrollándose de forma ininterrumpida mediante distintos proyectos de investigación, entre los que destacan los siguientes:

- “NEAR2050: future challenges for the rail sector” (H2020-S2RJU-OC-2015-01-2-73083), financiada por la convocatoria Shift 2 Rail de la Union Europea (2016).
- “SETA: an open, sustainable, ubiquitous data and service ecosystem for efficient, effective, safe, resilient mobility in metropolitan áreas” (H2020-ICT-2015-

688082), financiada por la commission Europea mediante la convocatoria H2020 (2016).

- “Modelización y control de la calidad percibida y ofrecida en tiempo real como instrumento para incentivar el uso del transporte público de viajeros (TRA2015-69903-R)”, financiada por el ministerio de economía y empresa, en colaboración de con los fondos FEDER de la unión europea (2016).
- Modelo científico de evaluación y mejora de la movilidad sostenible en Santander: definición de nuevas alternativas de aparcamiento (dynamic parking pricing) y evaluación y seguimiento de nuevo sistema de autobús de alto nivel de servicio (metro-tus), financiado por el servicio municipal de transportes urbanos de Santander (2017).
- Modeling the quality perception of public transport systems, financiado por Transportes Terrestres Cántabros, S.A.
- Movilidad compartida autónoma para las ciudades habitables del mañana - movi-city Financiado por la consejería de universidades e investigación, medio ambiente y política social del gobierno de cantabria (2019).

Como resultado del trabajo realizado en el grupo de investigación GIST desde su creación en 2004 se ha dotado a la Universidad de Cantabria de un Laboratorio de modelización dinámica del tráfico y transporte público, que permite la obtención de datos del estado del tráfico en la ciudad de Santander en tiempo real. Esta instalación ha sido posible gracias a la financiación recibida por parte del Ministerio de Ciencia, Innovación y Universidades del Gobierno de España, al igual que al apoyo recibido por parte del Ayuntamiento de Santander y del operador de transporte público TUS.

Fruto de la labor investigadora realizada en estos 13 años asociada a esta línea de investigación, se han conseguido publicar más de 50 artículos indexados en el Journal of Citation Report (JCR) de Thomson Reuters y se han defendido más de 100 ponencias en congresos nacionales e internacionales. Igualmente, han sido finalizadas con éxito un total de 3 tesis doctorales relacionadas con esta línea de investigación, todas ellas obteniendo la máxima calificación posible:

- “Metodología para el Estudio de la Demanda y la Calidad de los Sistemas Públicos de Bicicleta Mediante Datos Manuales y Automáticos”. Bordagaray, M. (2015).
- “Modelización de la interacción entre las variables de servicio y la demanda de viajeros de autobuses de tipo interurbano”. Rojo, M. (2011).
- “Modelización de la Calidad Percibida y Calidad Deseada del Transporte Público para Implantar Estrategias de Marketing en Áreas Urbanas”. Cecín, P. (2011).

La presente tesis, la cuarta dentro de la misma línea de investigación, está asociada principalmente al tercero de los proyectos listados anteriormente. Además, cabe destacar que el autor de esta tesis logró la concesión de una ayuda para la formación de profesorado universitario (FPU) para la realización de esta tesis, financiada inicialmente por el Ministerio de Educación y Formación Profesional y posteriormente por el Ministerio de Ciencia, Innovación y Universidades del Gobierno de España (Referencia FPU15/02990). Esta ayuda ha permitido a su vez, la obtención de una ayuda adicional para la financiación de una de las dos estancias breves realizadas en el Institute of Transport and Logistics

Studies (ITLS) de la University of Sydney (Academic Rank of World Unibersities: 68 General / 5 Tansportation Science & Technology). La otra estancia ha sido financiada con el apoyo del grupo de investigación en el que el doctorando ha realizado esta tesis y por los directores de tesis. El supervisor de ambas estancias ha sido el Dr. Chinh Ho (índice h: 10).

1.3 MOTIVACIÓN Y OBJETIVOS

Estudiar la calidad en el transporte público es una forma de definir las políticas enfocadas a mejorar el mismo. Un aspecto a analizar es la calidad percibida por los usuarios. Esta medida permite observar y cuantificar la satisfacción de los usuarios con respecto a un servicio de transporte público. La forma habitual de obtener esta información es mediante encuestas de satisfacción. Este tipo de encuestas se basan en preguntar a usuarios escogidos aleatoriamente cuál es su percepción con respecto a distintos aspectos que componen el servicio. Es habitual finalizar los cuestionarios preguntando la calificación general del servicio. Las encuestas requieren de un tiempo considerable para ser respondidas, con el consiguiente coste de tiempo y dinero. Por ello, aumentando la eficiencia de las encuestas se consigue un ahorro considerable para los operadores de transporte público.

Por lo tanto, el objetivo principal de la tesis ha sido estudiar la satisfacción de los usuarios en el transporte público optimizando la toma de datos sin perjudicar los resultados finales obtenidos. Para conseguirlo se ha analizó la literatura internacional para definir aquellas técnicas de encuesta que mejor se adaptasen al objetivo planteado. Gracias a la colaboración del servicio de transporte público de la ciudad de Santander se han podido realizar aplicaciones prácticas reales obteniendo los resultados que se muestran a lo largo de esta tesis.

La realización del objetivo principal se ha logrado gracias al cumplimiento de una serie de objetivos específicos que se citan a continuación:

- Definir un modelo que permita explicar la satisfacción de los clientes con un servicio de transporte público considerando la importancia de los distintos aspectos de un servicio y los distintos tipos de usuarios.
- Establecer un método que permita obtener los mismos resultados partiendo de una base de datos menor.
- Diseñar una metodología que permita obtener resultados de satisfacción e importancia de los distintos atributos que componen un servicio de forma más simplificada.
- Establecer una relación entre la importancia de las variables y el nivel de satisfacción
- Validar el nuevo método de encuesta propuesto con el método tradicional.

1.4 RELACIÓN ENTRE ARTÍCULOS DEL COMPENDIO

El compendio de artículos recogido en esta tesis se encuentra formado por cuatro artículos científicos indexados en el *Journal Citation Report (JCR)*. Estos artículos cumplen con los distintos objetivos específicos descritos en la tesis. Los artículos se encuentran englobados en una temática común, que es el estudio de la calidad percibida en el transporte público. Entendiendo como calidad percibida el estudio de la satisfacción y la importancia que muestran los clientes sobre distintos aspectos que conforman un servicio de transporte público. Los artículos presentados para el compendio son los siguientes.

- Artículo 1: Echaniz, E., dell'Olio, L., & Ibeas, Á. (2018). *Modelling perceived quality for urban public transport systems using weighted variables and random parameters*. *Transport Policy*, 67, 31-39.
- Artículo 2: Echaniz, E., Ho, C., Rodriguez, A., & dell'Olio, L. (2019). *Modelling user satisfaction in public transport systems considering missing information*. *Transportation*, 1-19.
- Artículo 3: Sañudo, R., Echaniz, E., Alonso, B., & Cordera, R. (2019). *Addressing the Importance of Service Attributes in Railways*. *Sustainability*, 11(12), 3411.
- Artículo 4: Echaniz, E., Ho, C. Q., Rodriguez, A., & dell'Olio, L. (2019). *Comparing best-worst and ordered logit approaches for user satisfaction in transit services*. *Transportation Research Part A: Policy and Practice*, 130, 752-769.

La investigación que ha dado lugar a la tesis ha seguido una evolución metodológica cronológica que se ve reflejada en el orden de los artículos, siendo el artículo 1 el más antiguo y el artículo 4 el más recientemente desarrollado. Los artículos se fundamentan principalmente en el estudio del transporte público de la ciudad de Santander (artículo 1, 2 y 4), sin embargo, en el artículo 3 se aumenta el alcance de la aplicación práctica a nivel europeo. A su vez, mientras que las aplicaciones realizadas en Santander se centran en el transporte público realizado mediante autobús, el artículo 3 se enfoca en el ferrocarril. Esta variación de ámbito de estudio es el resultado de la participación del autor de la tesis en el proyecto europeo NEAR2050 de la convocatoria Shift2Rail, perteneciente a la Unión Europea, donde el objetivo principal era el estudio de las necesidades de los clientes en el ferrocarril para horizontes temporales futuros.

Por lo tanto, el proceso seguido para el desarrollo de la tesis ha sido el siguiente. En primer lugar (artículo 1) se desarrolló un modelo matemático donde se incluía la importancia y la satisfacción de distintos atributos del transporte público de Santander. Mediante este modelo es posible estimar la satisfacción general con el servicio dependiendo del nivel de satisfacción de los distintos atributos. De este primer artículo se concluyó, entre otros aspectos, que la obtención de datos para el desarrollo de modelos complejos era costoso y debía ser optimizado. La optimización para la obtención de datos se planteó de dos formas distintas. Por una parte, se planteó la posibilidad de reducir el número de preguntas realizadas en cada encuesta e inferir esos datos no recogidos mediante algún método matemático (Artículo 2). Por otra parte, se consideró cambiar el tipo de pregunta realizada, de un tipo basado en evaluación tradicional (escalas Likert), a un tipo de preguntas más eficientes tipo Best-Worst. Este método se validó en el artículo

3. Por último, debido a que la aplicación realizada en el artículo 3 no se podía extrapolar directamente al ámbito de estudio de la ciudad de Santander, se realizó una encuesta que permitía comparar los resultados obtenidos mediante los dos métodos de encuestado en un mismo caso, siendo los resultados comparados en el artículo 4.

A continuación, se resumen brevemente los artículos empleados para constitución del compendio de acuerdo al orden cronológico de publicación, que a su vez es representativo de la evolución de la investigación.

El primer artículo (*Modelling perceived quality for urban public transport systems using weighted variables and random parameters*) fue aceptado para publicación el 31 de mayo de 2017 y finalmente publicado el 15 de Septiembre de 2018. En este artículo se propone una metodología que considera la importancia de los distintos atributos que componen un servicio de transporte público a la hora de realizar un estudio de calidad percibida. Los atributos son agrupados en distintos grupos, cada atributo es ordenado en base a su importancia dentro del grupo al que pertenece y a continuación los grupos son ordenados de mayor a menor importancia. Se proponen una serie de modelos ordered probit donde se incluyen, además, las variaciones sistemáticas y aleatorias en los gustos de los usuarios. La importancia de las variables se incluye en el modelo mediante una ponderación realizada por un factor que aumenta de valor de acuerdo a la importancia. El artículo concluye que a medida que la complejidad de los modelos aumenta, su capacidad de representar la realidad también. Sin embargo, el esfuerzo requerido para la obtención de los datos requeridos para la estimación de los modelos más complejos es demasiado alto en comparación a las mejoras de predicción obtenidas en los modelos.

El segundo artículo (*Modelling user satisfaction in public transport systems considering missing information*) fue aceptado para su publicación el 18 de marzo de 2019 y publicada online el 28 del mismo mes. Debido a que las encuestas de satisfacción pueden llegar a ser demasiado largas, en este artículo se analiza la posibilidad de obtener los mismos resultados de modelización partiendo de una base de datos menor. El principal problema a la hora de realizar encuestas de satisfacción en el transporte público es la duración de las mismas. Si las encuestas son demasiado largas los encuestados no disponen del tiempo suficiente para completarlas y los encuestadores deben abandonar el vehículo para acabar la encuesta o dejar la encuesta sin terminar. En ambos casos, la eficiencia del proceso de encuestado se ve reducida. Por lo tanto, en este artículo se plantea la hipótesis de obtener solo la mitad de los datos, simulando una situación donde solo se hicieran la mitad de las preguntas a cada usuario. Partiendo de esta hipótesis se plantea un proceso comparativo donde distintos modelos ordered probit son calculados basados en métodos de imputación. Método que se basa en completar las bases de datos partiendo de la información disponible. La comparativa se realiza entre tres modelos distintos, cada uno basado en un modo de imputación distinto: basado en modas, imputación única con modelos predictivos e imputación múltiple. Los resultados muestran que los modelos basados en imputación múltiple se comportan de forma similar a los modelos obtenidos con todos los datos. Concluyendo que al aplicar este método es posible reducir la longitud de las encuestas de satisfacción convencionales.

El tercer artículo del compendio (*Addressing the Importance of Service Attributes in Railways*) fue aceptado el 19 junio de 2019 y publicado el 21 de Junio de 2019. Este

artículo presenta un método cualitativo y cuantitativo combinado para la obtención de la importancia de los distintos atributos que componen un sistema de ferrocarril. La fase cualitativa se basa en proceso de participación mediante grupos focales y entrevistas en profundidad. Mediante estos métodos es posible definir la lista de atributos que mejor representa a un sistema de transporte ferroviario. La fase cuantitativa se centra en la realización de una encuesta de tipo Best-Worst (Caso 1) que permite establecer el nivel de importancia de cada atributo evaluado. Los resultados muestran que este tipo de encuestas permite definir la importancia de un gran número de variables de forma rápida y sin demasiada carga cognitiva para el encuestado. Obteniendo que para el caso del ferrocarril el precio, el tiempo de viaje y la intermodalidad son las variables más importantes, mientras que, los servicios adicionales no esenciales son los menos importantes. A su vez, se ha observado que el nivel de importancia de un atributo varía dependiendo de la frecuencia de uso del transporte del encuestado. Concluyendo que es necesario mejorar la intermodalidad del ferrocarril para atraer nuevos usuarios.

El cuarto y último artículo de los que componen el compendio (*Comparing best-worst and ordered logit approaches for user satisfaction in transit services*) fue aceptado para su publicación el 20 de octubre de 2019 y finalmente publicada online el 31 de octubre de 2019. En este artículo se analiza la posibilidad de reemplazar el método tradicional empleado para las encuestas de satisfacción por el método Best-Worst (caso 1), que se ha demostrado ser más eficiente y fácil de realizar. Para ello se ha utilizado una nueva encuesta realizada en la ciudad de Santander. Los resultados muestran que los niveles de satisfacción obtenidos para cada atributo mediante los dos métodos son notablemente similares. La importancia por otro lado difiere en ambos métodos, siendo los resultados obtenidos mediante Best-Worst más consistentes con la literatura existente. Además, se estima un modelo de regresión que permite obtener los resultados tradicionales mediante los datos Best-Worst con un margen de error pequeño.

1.5 ESTRUCTURA DE LA TESIS

Esta tesis se compone de un total de 7 capítulos. En este primer capítulo se ha realizado una introducción al conjunto de la tesis. En ella se concreta la normativa en el cual la tesis se encuentra enmarcada, los antecedentes existentes previos a la realización a la tesis y los cuales sirven para definir la motivación y los objetivos del proyecto de la tesis. Se explica a su vez, el hilo conductor existente entre los distintos artículos publicados a lo largo de la tesis y que componen el compendio de artículos.

El segundo capítulo muestra de forma resumida el análisis de la literatura internacional existente, hasta la fecha de realización de la tesis, en el ámbito de estudio del mismo.

El tercer capítulo resume la metodología empleada en los distintos artículos publicados y que se presentan como compendio en esta tesis. Al final de este capítulo se establece la relación entre los métodos explicados y cada uno de los artículos del compendio.

En el cuarto capítulo se recoge una transcripción literal de los 4 artículos que componen el compendio de artículos, así como sus principales indicadores bibliométricos de calidad.

En el capítulo 5 se discuten los principales resultados obtenidos en los artículos.

En sexto capítulo se establecen las conclusiones generales de la tesis, las cuales resumen y sintetizan las conclusiones específicas obtenidas para cada uno de los artículos. De la misma forma, basándose en estas conclusiones se establecen las futuras líneas de investigación.

El séptimo y último capítulo está compuesto por una traducción al inglés de las conclusiones y futuras líneas de investigación expuestas en el capítulo 6. Requisito obligatorio para la obtención de la mención internacional.

“Cuanto más sepáis de todo, mejor. Por saber muchísimo no os va a pasar nada malo, luego ya veremos.”

“The more you know about everything, the better. For knowing so much nothing bad will happen to you, then we'll see.”

(Rajoy, M., 2013)

CAPÍTULO 2: SÍNTESIS DEL ESTADO DEL ARTE

21 CALIDAD EN EL TRANSPORTE PÚBLICO

El estudio de la calidad percibida por los usuarios del transporte público (TP) es un requisito moderno. Los primeros estudios centrados en el concepto de la calidad percibida y la satisfacción de los clientes en el transporte fueron los de Parasuraman et al. (1988, 1985). Estos estudios consideraban los puntos de vista tanto de los operadores como de los usuarios, con el objetivo de comparar las expectativas de los usuarios con los objetivos de los operadores de transporte público.

Diversos estudios (Hensher et al., 2003; Hensher and Prioni, 2002; Mokonyama and Venter, 2013) han demostrado que la demanda asociada a un medio de transporte está directamente ligada a la satisfacción de los usuarios. Debido a esto, se han desarrollado contratos de gestión de transporte público basados en el rendimiento de los operadores.

Sin embargo, la mejora continua de un servicio de TP no es un ciclo sin fin. Cuando la calidad de un servicio y la satisfacción de los clientes es alta, es posible que una mejora en el servicio no genere un aumento en el número de usuarios. Friman (2004) estableció que la satisfacción de los usuarios no aumenta constantemente y que en efecto, este tiene un límite. Por otra parte, observando mejoras realizadas en un sistema de transporte, demostró que las mejoras realizadas no siempre eran percibidas por los usuarios de forma positiva, y que algunas de estas modificaciones no mejoraban los niveles de satisfacción global del sistema. Como consecuencia de este estudio se demuestra que para definir la satisfacción de los usuarios es necesario definir aquellos atributos del sistema que más influyen en la percepción de la calidad. Dicho de otra forma, es necesario definir tanto la satisfacción como la importancia de los atributos para realmente entender la calidad percibida en un sistema de TP.

Actualmente en Europa la calidad en el transporte público se encuentra definida por las normas EN 13816:2002 y la EN 15140:2006 (en España UNE-EN 13816 y UNE-EN 15140). La primera de ellas define el marco teórico y práctico de la calidad mientras que la segunda está más enfocada a las empresas operadoras de los servicios.

La norma EN13816 se desarrolló gracias a los resultados obtenidos en el proyecto europeo QUATTRO (EC, 1999) donde se analizaron distintos sistemas de TP de varios países de Europa. La aportación principal de esta norma reside en la definición del ciclo de calidad que gobierna el TP de pasajeros. Este ciclo se compone de 4 partes (Calidad objetivo, Calidad producida, calidad percibida y calidad esperada) que interactúan entre ellas. La Figura 1 muestra el esquema del ciclo de la calidad.

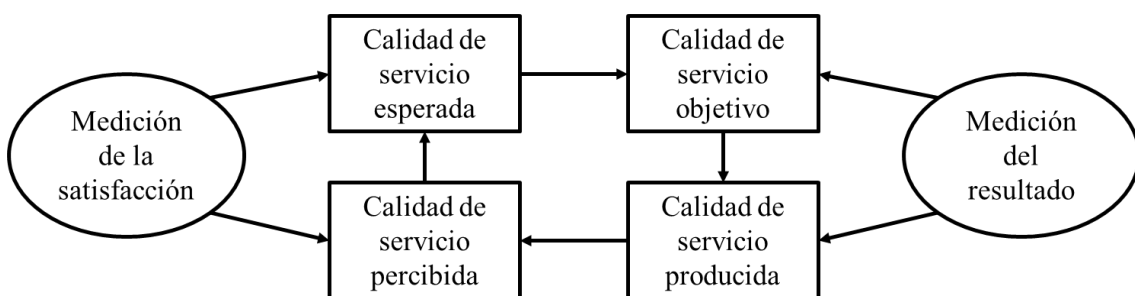


Figura 1 Ciclo de la Calidad de Servicio. Fuente: Aenor - UNE-EN 13816

La calidad esperada se define como la calidad que es requerida por los usuarios antes de utilizar un sistema de TP. La calidad objetivo es la definida en el proyecto, el cual, debe reflejarse en el contrato de gestión del servicio de TP y busca cumplir con la calidad esperada por los usuarios. La calidad ofrecida o producida es la calidad real conseguida por el operador. Por último, la calidad percibida es cómo los usuarios perciben la calidad producida por el operador. Se define como satisfacción de los usuarios la diferencia entre la calidad esperada y la calidad percibida. Por esto, en muchos casos los estudios de calidad percibida se encuentran referidos como estudios de satisfacción. Esta tesis se centra en analizar la calidad percibida por los usuarios optimizando la obtención de datos y el análisis de los mismos mediante distintas metodologías.

22 ENCUESTAS DE SATISFACCIÓN

Para poder recoger y analizar la calidad percibida por los clientes es necesario realizar encuestas, también llamadas encuestas de satisfacción. La forma habitual de realizar estas encuestas es mediante entrevistas presenciales (Bordagaray et al., 2014; dell’Olio et al., 2011, 2010; Rojo et al., 2013). Sin embargo, en los últimos años se han extendido métodos menos costosos para realizar las encuestas, como es el caso de Guirao et al. (2015), donde un cuestionario era proporcionado a los usuarios mediante el uso de códigos QR en las paradas. O los casos, por ejemplo St-Louis et al. (2014) donde la encuesta se realizó mediante un cuestionario online, cuya difusión se realizó mediante correos electrónicos o redes sociales. Otra forma de realización de encuestas es mediante llamadas telefónicas, sin embargo, este método no es muy común en encuestas de satisfacción de TP existiendo pocos estudios recientes que lo utilicen. Un ejemplo es el estudio realizado por Abenoza et al. (2017), donde se utilizó una base de datos muy extensa del barómetro sueco de transporte público compuesta por 450.000 encuestas telefónicas útiles realizadas entre los años 2001 y 2014. Este tipo de encuestas están siendo sustituidas por encuestas online por su menor coste de realización.

El proceso de obtención de datos es esencial para poder realizar un buen estudio de calidad percibida. Un aspecto importante a la hora de obtener una buena cantidad de respuestas con buena calidad es la duración requerida para completar la encuesta. Una encuesta larga permite obtener una gran cantidad de datos de cada individuo que permiten enriquecer el posterior análisis, no obstante, reduce en gran medida la tasa de encuestas completadas o respondidas, derivando en un menor número de observaciones útiles para un mismo presupuesto. En cambio, las encuestas más cortas permiten obtener un mayor número de observaciones en detrimento de la información obtenida en cada una de ellas. Por esta razón, es posible que los modelos estimados mediante el uso de estos datos pueden no ser del todo fiables al no disponer de todas las variables que afectan a la satisfacción de los usuarios. El equilibrio entre la longitud de la encuesta y el límite presupuestario depende en gran medida de la duración de las encuestas, el tamaño de la muestra necesaria y el método de encuesta escogido (presencial, app, online, ...).

La duración de una encuesta está directamente relacionada con el número de preguntas que se realizan en ella. Puesto que el objetivo de las encuestas es evaluar un servicio de TP al igual que los distintos aspectos que lo conforman, es necesario definir un conjunto

de atributos que representes de forma más completa posible el servicio de TP a evaluar. A lo largo del tiempo se han realizado varios estudios centrados en estos aspectos. Destacan especialmente los trabajos de Parasuraman et al. (1988) y Hensher et al. (2003) e igualmente el proyecto QUATTRO (EC, 1999). Parasuraman et al. (1988) definió la escala genérica SERVQUAL compuesto por un total de 22 atributos que permiten evaluar la calidad en distintos servicios. Hensher et al. (2003) desarrollaron un inventario de atributos enfocados en el análisis de la calidad en servicios de transporte público de pasajeros por carretera. Este inventario estaba compuesto por 13 atributos: Tiempo de viaje, precio del billete, tipo de billete, frecuencia, puntualidad, tiempo de acceso, disponibilidad de asientos en el autobús, información en la parada, tipo de acceso al autobús, instalaciones de la parada, temperatura a bordo, actitud del conductor y nivel de limpieza a bordo. El estudio más extensivo en cuanto al desarrollo de un listado de atributos que definen un servicio de TP fue realizado por el proyecto QUATTRO (EC, 1999), donde se definieron un total de 8 grupos (Servicio ofertado, Accesibilidad, Información, Tiempo, Atención al cliente, Confort, Seguridad, Impacto medioambiental) compuesto por un total de 99 atributos evaluables. La norma UNE-EN 13816 recoge estos mismos atributos a la hora de establecer un criterio de calidad a un servicio de TP, sin que sea necesario evaluar la totalidad de los mismos. Estos estudios son la base esencial utilizada actualmente, aunque requieren de modificaciones para adaptarse a los distintos servicios de TP y las características específicas de cada municipio o región. Varios autores han utilizado estas bases adaptadas para sus estudios de calidad percibida (de Oña et al., 2012; del Castillo and Benitez, 2012; Eboli and Mazzula, 2009; Eboli and Mazzulla, 2012; Kim and Chung, 2016; Metri, 2006). Una forma común de esta completar las variables más influyentes en un servicio de transporte público es mediante la realización de grupos focales con los distintos usuarios del servicio (Ibeas et al., 2011), lo cual permite establecer las características propias de un sistema concreto.

Por último, la mayoría de los estudios no se limitan a obtener solo los datos relacionados con la calidad percibida. Es una práctica habitual recolectar información socioeconómica de los encuestados (Abenoza et al., 2017; de Oña et al., 2016d; Diana, 2012; Filipović et al., 2009; Joewono and Kubota, 2007; A. Minhans et al., 2015; Susilo and Cats, 2014). Mediante estos datos es posible definir el tipo de usuario que utiliza el sistema de transporte y sus necesidades. Cada ciudad o región tiene sus propias características y tradiciones, por lo tanto, las variables socioeconómicas escogidas para definir a los usuarios también varían. Por lo general, las variables más utilizadas son: género, edad, situación laboral, nivel de educación o motivo del viaje. Además, de poder caracterizar el tipo de usuario, diversos estudios (Bordagaray et al., 2014; de Oña et al., 2016d; dell'Olio et al., 2010) han demostrado que la satisfacción de los clientes cambia con respecto a sus características socioeconómicas. Por lo tanto, la recolección de este tipo de datos gana mayor importancia a la hora de realizar el posterior análisis y modelización.

Una vez definidas las variables o preguntas que se incluirán en la encuesta, es necesario establecer una escala de valoración que permita obtener la información requerida. Establecer un tipo de pregunta u otro hará que la complejidad de la encuesta varíe. En las encuestas de satisfacción, por lo general, se utilizan preguntas con respuestas cerradas. En el caso de las variables socioeconómicas, es habitual proponer opciones de respuesta que permitan obtener todas las opciones posibles. En cuanto a las evaluaciones de

satisfacción, las posibles respuestas se suelen presentar como escalas de opciones, existiendo la posibilidad de establecer una escala cuantitativa o cualitativa. El uso de escalas cuantitativas (de Oña et al., 2013; de Oña and de Oña, 2015; dell’Olio et al., 2017) requiere de proporcionar las opciones de respuesta en base a valores numéricos, siendo habituales valores comprendidos entre 0 y 5 o 0 y 10, siendo por lo general el valor mayor la opción más positiva. En cuanto a las preguntas cualitativas (Bordagaray et al., 2014; dell’Olio et al., 2010; Gonzalo-Orden et al., 2011; Rojo et al., 2013), las opciones de respuestas suelen seguir un tipo de escala tipo Likert (Likert et al., 1934), donde se definen varios niveles (por lo general no superior a 5) y se definen las distintas opciones desde “muy satisfecho” a “muy insatisfecho” o similar. Es común asociar los valores cualitativos a valores numéricos en el análisis de los datos o en la modelización (Alonso et al., 2018; dell’Olio et al., 2010; Hensher et al., 2010; Rose and Hensher, 2018). Dentro de una encuesta se suele utilizar una misma escala tanto para la satisfacción como para la importancia. Existen otros métodos de encuesta utilizados para este tipo de análisis, aunque su uso es más limitado. Este es el caso de Beck and Rose (2016) que utilizaron un tipo de pregunta BW (Louiére et al., 2015) para establecer tanto la satisfacción, como la importancia de distintos atributos, resultados que compararon con una escala cualitativa más tradicional. Los resultados mostraron que las encuestas de tipo BW permitían captar mejor las correlaciones existentes entre la satisfacción y la importancia de los atributos. La elección del tipo de pregunta estará condicionada por el método de análisis que se desea utilizar a posteriori, puesto que los datos obtenidos en las encuestas deben satisfacer los requisitos necesarios para la modelización.

Establecer la importancia de los distintos atributos es otro aspecto muy común en las encuestas de satisfacción. Sin embargo, el estudio de la importancia no siempre se realiza de forma directa mediante preguntas explícitas. (Cao and Cao, 2017) realiza una buena explicación de las dos formas de establecer la importancia de los atributos. El primer enfoque reside en obtener las respuestas de forma explícita (importancia explícita). Esto es, los encuestados responden a una pregunta similar a la de la satisfacción, pero centrándose en la importancia. Las repuestas pueden realizarse mediante las mismas escalas definidas para la satisfacción. El segundo enfoque es la importancia derivada del análisis de la satisfacción (importancia implícita). La importancia implícita se obtiene relacionando la satisfacción de los atributos con la satisfacción global del servicio. Esto se realiza mediante algún método de análisis o modelización, obteniendo la importancia estadística inferida de los modelos. Los distintos métodos utilizados en la literatura para este análisis son expuestos en el siguiente apartado. Un gran número de estudios han utilizado el método de la importancia explícita para obtener la información requerida para el análisis (Beck and Rose, 2016; Guirao et al., 2016; Rose and Hensher, 2018). Sin embargo, la inclusión de estas preguntas en los cuestionarios genera un aumento del tiempo de encuestad La eficiencia del proceso se puede mejorar utilizando métodos de encuesta alternativos, como es el caso de Beck and Rose (2016), donde se utiliza el tipo de encuesta BW para obtener la importancia de los distintos atributos. En cuanto a la importancia implícita, a lo largo de los años se han utilizado distintos métodos de modelización que han permitido obtener distintos resultados. Existe una variedad considerable de estudios (Allen et al., 2019, 2018b; Bordagaray et al., 2014; de Oña and de Oña, 2015; dell’Olio et al., 2018; Román et al., 2014) con metodologías diferentes. Sin embargo, de forma general los resultados obtenidos son realmente similares. Siendo

las variables más importantes en un sistema de transporte la frecuencia, la fiabilidad y el tiempo de viaje. En menor medida también son considerados de cierta importancia la suavidad de conducción y el confort de los vehículos. Aquellos estudios centrados más en analizar la afección de distintos atributos en la demanda de pasajeros (Ahern and Tapley, 2008; Brons et al., 2009; dell'Olio et al., 2018; Givoni and Rietveld, 2007; Liu et al., 2016; Park and Ha, 2006; Román et al., 2010) muestran además de la importancia de los atributos anteriormente mencionados el precio del billete y los tiempos de acceso/egreso a las paradas o estaciones en el caso del transporte ferroviario.

Un resultando interesante fue el obtenido por Cao and Cao (2017), donde se compararon los resultados obtenidos para un mismo caso de la importancia explícita e implícita, y se observó que los resultados de ambos resultados eran distintos.

Las pautas mencionadas en los párrafos anteriores son generales y de aplicación en la práctica totalidad de los estudios que se analizan en esta tesis. Cada estudio muestra sus propias características y en cada caso desarrolla una explicación detallada de la encuesta o encuestas realizadas. Estos estudios definen claramente el método de encuestado, la población objetivo y la muestra obtenida, el diseño de la encuesta y el tipo de resultados obtenidos.

23 ENFOQUE METODOLÓGICO DEL ANÁLISIS DE LA CALIDAD PERCIBIDA

Los métodos de análisis y modelado escogidos para cada estudio condicionan el diseño y la realización de las encuestas de satisfacción. En la literatura pueden observarse una gran variedad de métodos que han permitido analizar la satisfacción en distintos modos de transporte de forma exitosa.

El método más sencillo y que se encuentra generalmente incluido en la mayoría de los estudios es el análisis estadístico básico de los resultados. Calculando las medias y las variaciones de los resultados permiten establecer una primera imagen del servicio que se está estudiando.

Como se ha mencionado anteriormente, los métodos de modelado permiten establecer el peso que tienen los distintos atributos de un servicio de transporte público a la hora de explicar la satisfacción general de los usuarios. Dicho de otra forma, permite definir mediante un enfoque matemático la importancia que cada aspecto del servicio tiene para el usuario, esto es, la importancia implícita.

A lo largo de los años los métodos utilizados han sido muy diversos. En los estudios de (de Oña and de Oña, 2015; dell'Olio et al., 2017) se muestran una extensa revisión de los métodos utilizados para este fin.

El método más sencillo y popular es el análisis descriptivo de los datos. Donde el promedio y la desviación de los niveles de satisfacción de los distintos atributos y del servicio en su conjunto se utiliza para representar la satisfacción de los usuarios. Existe una gran variedad de estudios que aplican este método (de Oña et al., 2013; Eboli and Mazzula, 2009; Eboli and Mazzulla, 2010, 2007). Además, este es un método utilizado

en la prácticamente totalidad de los estudios de satisfacción actuales como una primera aproximación para posteriormente aplicar métodos más avanzados.

Los métodos más comunes utilizados en la última década han sido las ecuaciones estructurales y los modelos de elección discreta. En el uso de ecuaciones estructurales destacan los trabajos de Allen et al. (2019), Eboli and Mazzulla (2015), Hapsari et al. (2016), Machado et al. (2018), Rose and Hensher (2018) o Shen et al. (2016). Este método ha mostrado ser muy útil para entender los distintos mecanismos que explican la satisfacción de los usuarios. Las ecuaciones estructurales se han aplicado en el caso del transporte público por carretera (Allen et al., 2018b; Das et al., 2017; de Oña et al., 2016b, 2013; Eboli and Mazzulla, 2007), como en otros medios de transporte (Eboli and Mazzulla, 2015; Machado et al., 2018; Rose and Hensher, 2018; Shen et al., 2016; Tam et al., 2010). En todos los casos, la aplicación de este método ha demostrado ser útil para la obtención de resultados coherentes e interesantes. Por otro lado, la metodología basada en modelos de elección discreta, más concretamente los modelos de dato ordenados, han sido ampliamente utilizados en el ámbito de la calidad en el transporte público. Destacan los trabajos de (Bordagaray et al., 2014; dell'Olio et al., 2010; dell'Olio et al., 2017; Hensher et al., 2010; Rojo et al., 2013; Tyrinopoulos and Antoniou, 2015). Una de las ventajas de los modelos de datos ordenados es la posibilidad de captar la no-linealidad entre las distintas opciones de elección. Un hecho que se corresponde con la realidad de las aplicaciones, puesto que se ha demostrado que es más costoso mejorar la calidad percibida hacia un servicio bueno que mejorar un servicio en un inicio malo o deficiente. Otra forma basada en un modelo de elección discreta ha sido la realización de encuestas de preferencias declaradas (SP), que muestran al encuestado una serie de tareas de elección y le piden que elija la que más le gusta. Los datos se analizan utilizando algún tipo de modelo de elección discreta. Por ejemplo, Román y otros (2014) utilizaron el modelo logit multinomial (MNL) y el modelo logit mixto (ML) para examinar los servicios de transporte público en Gran Canaria (España). Del mismo modo, dell'Olio et al. (2011) utilizaron un modelo MNL para analizar la calidad deseada por los futuros usuarios.

Si bien estos métodos aglutinan una gran variedad de casos de estudio, en la literatura existen otros métodos que también han demostrado ser fiables y útiles para el análisis de la calidad. Entre otros se puede encontrar el uso de árboles de decisión (de Oña et al., 2016d, 2014; Hernandez et al., 2016; Tsami and Nathanail, 2017), análisis multicriterio (Hrncir et al., 2017; Nathanail, 2008), análisis factorial (Burlando et al., 2016; Cats et al., 2015; Wen et al., 2005), métodos basados en redes neuronales (de Oña et al., 2017; dell'Olio et al., 2017; Garrido et al., 2014) o métodos de regresión múltiple (Das et al., 2017; Guirao et al., 2016; Mouwen, 2015).

Indiferentemente del método de análisis utilizado, los resultados de los estudios son generalmente bastante similares. Uno de los aspectos en los que coinciden los distintos métodos, es que, la muestra analizada es heterogénea y debe ser analizada de forma que se consiga captar esa variabilidad en los gustos. Una forma de incluir la heterogeneidad de los usuarios es mediante el uso de clusters (de Oña et al., 2016a, 2016d; Vicente and Reis, 2016), lo cual permite establecer distintos tipos de usuarios clasificados por grupos de acuerdo a sus características. Otra forma de incluir esa variabilidad en los gustos de los usuarios es mediante la inclusión de interacciones en los modelos de datos ordenados

(Ibeas et al., 2014). Estas interacciones están basadas en las variables socioeconómicas incluidas en la encuesta, y sirven para explicar la variación sistemática o medible de la muestra. Si la muestra analizada es heterogénea pero no se disponen de los datos o la información necesaria para explicar esa heterogeneidad es posible incluir una variabilidad a los modelos mediante la inclusión de parámetros aleatorios (William H. Greene and Hensher, 2010; Hensher and Mulley, 2014). De esta forma se establece que los parámetros que se estiman en los modelos tendrán asociado una variabilidad con un origen desconocido pero que se distribuirá de acuerdo a una función estadística conocida.

Una de los requerimientos de los modelos de datos ordenados es la necesidad de disponer de una base de datos completa que permita estimar los parámetros asociados al modelo. Cuando no se dispone de la base de datos completa es necesario imputar aquellos datos faltantes. Dicho de otra forma, completar la base de datos con la información disponible.

El cálculo estadístico relacionado con los datos faltantes ha evolucionado considerablemente en los últimos años. El estudio realizado por Schafer and Graham (2002) proporciona una buena visión de los tipos de datos faltantes que pueden aparecer y los distintos métodos de imputación que pueden aplicarse. De acuerdo con la clasificación realizada por Rubin (1976), dependiendo de la naturaleza por la cual falta la información, es posible establecer distintos tipos de información faltante. La información faltante se define como MAR (Missing at Random) cuando la distribución asociada a la falta de datos no depende de los mismos datos faltantes. Sin embargo, este caso sí que permite que la distribución de la falta de datos dependa de los datos observados. Por esta razón, se define un caso especial dentro de MAR que se define como MCAR (Missing completely at random). Lo que significa que la razón de la falta de datos no está asociada ni a los datos observados ni a los datos faltantes. Si se incumple la condición de MAR entonces la información faltante se define como MNAR (missing not at random).

En el caso asociado al segundo artículo del compendio presentado en esta tesis, la información faltante es deliberada pero establecida de forma aleatoria, siendo, por lo tanto, referido a la notación anterior, una información faltante de tipo MCAR. En Graham et al. (1996) este tipo de escenarios se definen como patrones de valores faltantes planeados (Planned missing value patterns). En otras palabras, la encuesta es intencionalmente planeada con información faltante. Cuando la información faltante es del tipo MCAR, es posible aplicar métodos de imputación simples (Donders et al., 2006). Por ejemplo, existe la posibilidad de eliminar aquellas observaciones con datos faltantes (listwise deletion). Este método puede aplicarse siempre y cuando el número de observaciones con datos faltantes no sea muy numeroso y no afecte a la robustez de la muestra. Dentro de estos métodos simples, otra posibilidad es la de reemplazar aquellos datos faltantes con las medias calculadas con los datos observados. Por último, existe la posibilidad de insertar los datos faltantes basándose en la imputación simple. El cual se basa en imputar los valores faltantes considerando la información disponible y otorgando un valor plausible. Sin embargo, aunque los métodos sencillos aportan resultados aceptables, varios estudios recomiendan la utilización de métodos más avanzados como son los métodos basados en máxima verosimilitud (Graham et al., 1996) o imputación múltiple (Donders et al., 2006; Graham et al., 2003, 1996). En el caso de Donders et al. (2006), se demuestra además que el método de imputación múltiple (MI) es especialmente recomendado para casos de MCAR. De la misma forma, Graham and Schafer (1999)

demonstraron que la imputación múltiple funciona bien en tamaños de muestra pequeños y con valores faltantes cercanos al 50%.

La imputación múltiple, desarrollada inicialmente por Rubin (1977), ha demostrado ser un método muy efectivo para obtener información faltante. Este método es muy popular en el ámbito de las ciencias sociales y medicina. Como ejemplo de su aplicación se encuentran: Alegria et al., 2004; Allison, 2000; Burton et al., 2007; König et al., 2018; Love et al., 2018; Newgard et al., 2018; Pettersson et al., 2018; Phan et al., 2016; Roth, 1994; Sterne et al., 2009; Troyanskaya et al., 2001; van Buuren et al., 1999 o Zou, 2015. Esta metodología ha tenido poca aplicación en el ámbito del transporte. Un uso particular ha sido la de completar los datos de tráfico asociados a los flujos o a los detectores (Chiou et al., 2014; Henrickson et al., 2015; Li et al., 2015; Tang et al., 2015).

Por ultimo cabe mencionar que la metodología Best –Worst (BW) aplicada en varios estudios de esta tesis requiere de una modelización ligeramente distinta a las utilizadas en los estudios mencionados anteriormente. A lo largo de los años han existido dos tendencias principales a la hora de estudiar las respuestas de tipo BW. Por una parte, se han utilizado métodos basados en puntuaciones, como son la forma analítica cerrada (Analytical closed form) (Lipovetsky and Conklin, 2014) o las puntuaciones BW normalizadas (Normalized BW scores) (Louviere et al., 2015). Por otra parte, la otra vertiente se asemeja más al tipo de modelización utilizada en los estudios de satisfacción, puesto que se basan en el uso de modelos de elección discreta. Este es el caso de los estudios de Beck and Rose (2016) o Marley and Pihlens (2012) por ejemplo. El modelo más utilizado es, al igual que en el caso de los estudios de encuesta PR, el modelo Logit Multinomial. En Marley et al. (2016) se realiza una comparativa de los tres métodos, los resultados muestran que los métodos basados en puntuación y los basados en modelos de elección discreta se encuentran muy relacionados entre sí.

“Esta es tu última oportunidad. Después, ya no podrás echarte atrás. Si tomas la pastilla azul, fin de la historia. Despertarás en tu cama y creerás lo que quieras creerte. Si tomas la roja, te quedas en el País de las Maravillas y yo te enseñaré hasta dónde llega la madriguera de conejos. Recuerda, lo único que te ofrezco es la verdad, Nada más.”

“This your last chance. After this, there is no turning back. You take the blue pill, the story ends. You wake up in your bed and believe whatever you want to. You take the red pill, you stay in Wonderland, and I show you how deep the rabbit hole goes. Remember, all I'm offering is the truth. Nothing more.”

Morpheus, The Matrix (1999)

CAPÍTULO 3: METODOLOGÍA EMPLEADA

3.1 MODELOS DE ELECCIÓN DISCRETA

Los modelos de elección discreta son modelos econométricos que sirven para definir la probabilidad de escoger una opción discreta de entre un grupo de alternativas. Su utilización en el mundo del transporte se encuentra muy extendido siendo habitual utilizarlo para la elección de modos de transporte o rutas. Este tipo de modelos se basan en la definición de la utilidad U_{nsj} asociada a cada alternativa de elección j , de entre un grupo de alternativas o situación de elección (escenario) s y percibido por un individuo n . La utilidad de una alternativa define en cierta forma lo atractivo que una opción resulta para cada persona. La utilidad se divide en dos componentes, una componente observable y medible, V_{nsj} , y una componente residual no observable, ε_{nsj} , de forma que:

$$U_{nsj} = V_{nsj} + \varepsilon_{nsj} \quad (1)$$

La componente observable es comúnmente asumida como una relación lineal entre los niveles de los atributos x de cada alternativa j y su correspondiente parámetro (peso) β .

$$U_{nsj} = \sum_{k=1}^K \beta_{nk} x_{nsjk} + \varepsilon_{nsj} \quad (2)$$

Donde β_{nk} representa la utilidad marginal o parámetro de peso asociado a un atributo k para el individuo n .

Para poder modelizar una elección es necesario realizar una serie de asunciones, siendo la más común la de establecer que el componente no observable, ε_{nsj} , se distribuye de acuerdo a una distribución de densidad aleatoria definida, $f(\varepsilon_{nsj})$. La elección específica de funciones de densidad conduce a modelos econométricos alternativos. Siendo los más comunes el modelo Probit (la componente no observable se distribuye de acuerdo a una distribución normal) y el modelo Logit (la componente no observable se distribuye de acuerdo a la distribución Gumbel de valores extremos).

De acuerdo con lo descrito anteriormente, los modelos de elección discreta permiten calcular la probabilidad con la que un individuo n en una situación de elección s escoja la opción i de entre un grupo de distintas alternativas j . Siendo la alternativa i escogida si su utilidad es la máxima de entre todas las alternativas:

$$P_{nsi} = \text{Prob}(U_{nsi} > U_{nsj}, \forall j \neq i) = \text{Prob}(V_{nsi} + \varepsilon_{nsi} > V_{nsj} + \varepsilon_{nsj}, \forall j \neq i) \quad (3)$$

Que puede también escribirse como:

$$P_{nsi} = \text{Prob}(\varepsilon_{nsi} - \varepsilon_{nsj} > V_{nsj} - V_{nsi}, \forall j \neq i) \quad (4)$$

Lo que refleja la probabilidad de que las diferencias en los términos aleatorios, $\varepsilon_{nsi} - \varepsilon_{nsj}$ sea menor que las diferencias entre las componentes observables de la utilidad, $V_{nsj} - V_{nsi}$.

3.1.1 Modelo Logit Multnomial

El modelo Logit se trata del modelo de elección discreta más comúnmente utilizado en la actualidad. Están basados en la asunción de que los efectos no observables del modelo se extraen de una distribución de valores extremos generalizada (GEV). Más concretamente asumiendo una distribución GEV de tipo 1 o distribución Gumbel.

Para poder estimar los modelos una serie de normalizaciones son necesarias. En primer lugar, es necesario definir los valores de la posición y la escala de la utilidad. Considerando la definición de la utilidad anteriormente descrita basada en una componente observable V_{nsj} y una componente no observable ε_{nsj} , se define la varianza del efecto no observable para la alternativa j como $Var(\varepsilon_{nsj}) = \sigma_j^2$. El valor de la varianza para una distribución GEV1 no estandarizada es $\sigma_j^2 = \pi^2/6\lambda_j^2$, donde λ_j es el parámetro de escala de la distribución. Debido a que los datos obtenidos para la modelización no permiten obtener el valor absoluto de la utilidad, sino que proporcionan un ranking entre las alternativas, no es posible estimar el parámetro de escala desconocido. Por lo tanto, es necesario normalizar la escala de la función de utilidad de la siguiente forma:

$$U_{nsj} = \lambda_j V_{nsj} + \varepsilon_{nsj} \quad (5)$$

A su vez, los modelo Logit se especifican considerando una restricción general de que los efectos no observables son iguales para todas las alternativas j . Para ello es necesario establecer algún tipo de normalización para σ_j^2 . Debido a que $\sigma_j^2 = \pi^2/6\lambda_j^2$, entonces es necesario establecer un parámetro de escala λ_j igual para todas las alternativas. Lo más habitual es establecer $\lambda_j = 1$, con lo que se consigue que $\sigma_j^2 = \frac{\pi^2}{6} = 1,6449$. Al establecer la normalización del parámetro de escala es equivalente a normalizar la escala de la función de utilidad.

Dependiendo del modelo Logit a estimar es necesario añadir una serie de restricciones adicionales. En el caso del Logit Multinomial (MNL), se considera que las covarianzas entre alternativas j son 0. Esto quiere decir que las alternativas son independientes entre sí.

Considerando las restricciones mencionadas hasta el momento, es habitual utilizar las siglas IID (Independently and Identically Distributed) para definir las restricciones asociadas a los modelo MNL (IID GEV1).

La mayor bondad del modelo MNL en comparación a modelos más complejos es la posibilidad de calcular las probabilidades de elección de una alternativa de forma sencilla mediante una formula cerrada, sin necesidad de simulaciones ni iteraciones. Definiéndose de la siguiente forma:

$$\text{Prob(Alt } j \text{ sea escogida)} = \frac{\exp(V_{nsj})}{\sum_{j=1}^J \exp(V_{nsj})}, j=1, \dots, J. \quad (6)$$

Por lo que la probabilidad de escoger la alternativa j de entre todas las alternativas J está definida por las diferencias entre las utilidades observables existentes entre las distintas alternativas.

3.1.2 Modelo Mixed Logit (Parámetros aleatorios)

Una de las limitaciones que tiene el modelo MNL, además las asociadas a las asunciones realizadas al componente aleatorio de la utilidad, reside en la definición de un peso β_k constante para todos los individuos encuestados. Siendo este un valor promedio asociado a cada variable para la muestra analizada. Por ello, esta homogeneidad en los parámetros estimados no permite establecer distinciones entre las distintas personas encuestadas, lo cual quiere decir que no se considera la heterogeneidad de la muestra.

La heterogeneidad existente entre las personas encuestadas puede definirse de dos maneras. Por una parte, es posible analizar las características socioeconómicas de los encuestados incluyendo preguntas adicionales en los cuestionarios. Estas variables pueden incluirse en los modelos siguiendo varios métodos, como puede ser, mediante interacciones. Estas interacciones representan las variaciones sistemáticas en los gustos de los usuarios. Por otra parte, es posible definir una heterogeneidad aleatoria en los modelos considerando que los parámetros estimados varían de un individuo a otro y esa variación está definida por una distribución estadística conocida. Este es el caso del modelo mixed logit, también conocido por el nombre de logit de parámetros aleatorios, kernel logit o mixed multinomial logit (MMNL). En este modelo, la probabilidad de escoger una alternativa dependerá de los parámetros aleatorios con distribuciones definidas por el analista. Siendo, por lo tanto:

$$\text{Prob}(\text{elección}_{ns} = j \mid \mathbf{x}_{nsj}, \mathbf{z}_n, \mathbf{v}_n) = \frac{\exp(V_{nsj})}{\sum_{j=1}^J \exp(V_{nsj})} \quad (7)$$

Con:

$$\begin{aligned} V_{nsj} &= \beta'_{nsj} \mathbf{x}_{nsj} \\ \beta_n &= \beta + \Delta \mathbf{z}_n + \Gamma \mathbf{v}_n \end{aligned} \quad (8)$$

Donde \mathbf{x}_{nsj} representa los valores de los atributos k de la alternativa j en una situación de elección s para el individuo n . \mathbf{z}_n es un conjunto de características M de cada individuo n que influyen en el valor de los parámetros. Finalmente, \mathbf{v}_n es un vector de K variables aleatorias con media cero y varianza conocida y ninguna covarianza.

El modelo de elección incorpora tanto la heterogeneidad observada como la no observada. La heterogeneidad observada se representa mediante el término $\Delta \mathbf{z}_n$, mientras que la heterogeneidad no observada se incluye mediante el término $\Gamma \mathbf{v}_n$. Los parámetros a estimar son el vector de constantes β , la matriz $K \times M$ de parámetros Δ y los elementos no nulos de la matriz triangular inferior de la descomposición de Cholesky Γ .

El modelo mixed expuesto puede modificarse para definir caso o modelos concretos. Por ejemplo, un modelo MNL con parámetros aleatorios puro se consigue cuando $\Delta = 0$ y Γ es diagonal. O un modelo MNL se definiría con $\Delta = 0$ y $\Gamma = 0$. Si bien, se elimina la

heterogeneidad observada existente en el modelo mixed, esta heterogeneidad puede añadirse al modelo mediante interacciones entre las variables socioeconómicas y los atributos k , consiguiendo de esta forma un modelo que considera la variación tanto sistemática como aleatorio en los gustos de los usuarios.

3.1.3 Modelo MNL para encuestas Best-Worst

Para modelizar los datos Best-Worst (BW) obtenidos en las encuestas, es necesario entender la diferencia existente con los modelos de elección discreta convencionales. Los modelos anteriormente descritos se basan en la elección de una alternativa como la opción más atractiva para el encuestado. Sin embargo, en las preguntas de tipo Best-Worst el usuario debe realizar dos elecciones, la mejor y la peor. Por ello, los modelos de elección no se basan en el cálculo de probabilidades para la elección de una alternativa, sino que establecen la probabilidad de escoger una alternativa mejor opción combinado con la probabilidad de escoger otra alternativa como peor.

Puesto que las preguntas BW se clasifican en tres casos distintos, tal y como se explica en el apartado 3.3.2, es necesario establecer un modelo distinto para cada caso. Para este apartado se ha utilizado como base el libro “*Best-Worst Scalling Theory, methods and application*” (Louviere et al., 2015).

La elección entre distintos perfiles (Caso 3) puede definirse como el caso genérico de modelización. Se definen un total de m atributos, siendo $M = \{1, \dots, m\}$. Se asume que el atributo i tiene $q(i)$ niveles. Por lo tanto, un perfil j se trata de un vector compuesto por m componentes donde cada componente i obtiene uno de los niveles $q(i)$ de ese componente. Siendo:

$$\mathbf{x}_j = (x_{j1}, \dots, x_{jm}) \quad (9)$$

Donde x_{ij} , $i = 1, \dots, m$, denota el nivel del atributo i en el perfil j . A su vez se define el vector β :

$$\beta = (\beta_1, \dots, \beta_m) \quad (10)$$

Que se trata del vector de coeficientes de utilidad de cada atributo i .

Las probabilidades de escoger un perfil i como mejor opción (opción best) entre un set de opciones X se denomina $P_B(i|X)$, respectivamente la opción worst se define como $P_W(i|X)$. La probabilidad conjunta de escoger un perfil i como la mejor y un perfil $i' \neq i$ como peor opción en un set X se denomina $P_{BW}(ii'|X)$.

El modelo MNL asume que la probabilidad de escoger el perfil i como mejor opción en un set W se define como:

$$P_B(i|X) = \frac{\exp \beta'_i \mathbf{x}_i}{\sum_{j \in X} \exp \beta'_j \mathbf{x}_j} \quad (11)$$

Respectivamente:

$$P_W(i'|X) = \frac{\exp(-\beta'x_i)}{\sum_{j \in X} \exp(-\beta'x_j)} \quad (12)$$

Por consiguiente se asume que el coeficiente de utilidad de cada atributo al escoger la opción worst es el negativo del coeficiente al escoger la opción best.

Estos modelos representan elecciones independientes de las opciones best y worst, si embargo la probabilidad conjunta debe satisfacer una restricción, no debe existir un opción i con $P_B(i|X) = P_W(i|X) = 1$, esto es, la mejor y la peor opción deben ser distintas. En consecuencia, la probabilidad de escoger el perfil i como mejor opción y el perfil $i' \neq i$ como peor opción de entre un set de alternativas X se define como:

$$P_{BW}(ii'|X) = \frac{\sum_{j \in X} \exp(\beta'(x_i - x_{i'}))}{\sum_{j' \neq j} \exp(\beta'(x_i - x_{j'}))} \quad (13)$$

Para el Caso 2, la elección en perfiles, el set X que se presenta es un único perfil x_k y las opciones de elección son los distintos niveles de los atributos x_{kr} , $r = 1, \dots, m$, de ese mismo perfil. Modificando la notación anterior para añadir esta restricción resulta en que la probabilidad de escoger un nivel de atributo x_{ki} como la mejor opción dentro de un perfil x_k y el atributo $x_{ki'}$, $i' \neq i$, como peor opción es:

$$P_{BW}(ii'|x_k) = \frac{\exp(\beta_i x_{ki} - \beta_{i'} x_{ki'})}{\sum_{j' \neq j} \frac{\exp(\beta_j x_{kj} - \beta_{j'} x_{kj'})}{\exp(\beta_j x_{kj} - \beta_{j'} x_{kj'})}} \quad (14)$$

En el último caso, el Caso 1, la elección entre objetos, el set X que se presenta no contempla una estructura multi-atributo de distintos niveles ni perfiles. Por lo tanto, se sustituye la notación x_k por un objeto genérico k . De modo que la probabilidad de escoger el objeto i como mejor opción y el objeto $i' \neq i$ como peor opción en un set de elección X se define de la siguiente forma:

$$P_{BW}(ii'|X) = \frac{\sum_{j \in X} \exp[u(i) - u(i')]}{\sum_{j' \neq j} \exp[u(j) - u(j')]} \quad (15)$$

Donde u representa el valor de la utilidad de cada atributo a estimar.

3.2 MODELOS DE DATOS ORDENADOS

Los modelos de datos ordenados, en su forma más contemporánea, basados en regresión, fueron propuestos por McKelvey and Zavoina (1975, 1971). El objetivo de estos modelos es analizar datos ordenados, categóricos y elecciones no cuantitativas.

La teoría en la que se basan estos modelos reside en dividir una utilidad latente continua, variable q^* , que es observada de forma discreta mediante un sistema de límites que acota dicha variable en franjas de elección.

Se establece un modelo de regresión latente, donde la variable dependiente q_i^* es definida mediante una regresión compuesta por unas variables sistemáticas observables v_i que multiplican a las utilidades marginales θ' y una componente aleatoria desconocida ε_i .

$$q_i^* = \theta'v_i + \varepsilon_i, i = 1, \dots, n \quad (16)$$

La variable latente es dividida en distintas bandas J mediante $J + 2$ parámetros de límite μ . La variable latente discreta se define como q_i . Tanto las utilidades marginales como los parámetros de límite deben de estimarse utilizando n observaciones.

$$\begin{aligned} q_i &= 0 \text{ if } \mu_{-1} < q_i^* \leq \mu_0, \\ &= 1 \text{ if } \mu_0 < q_i^* \leq \mu_1, \\ &= 2 \text{ if } \mu_1 < q_i^* \leq \mu_2, \\ &= \dots \\ &= J \text{ if } \mu_{J-1} < q_i^* \leq \mu_J \end{aligned} \quad (17)$$

La idea clave del modelo es que no se trata de una acumulación de observaciones discretas que han de ser ordenadas de alguna forma, sino de una transformación de una única variable continua que ha de ser ordenada.

Se asume que el término aleatorio ε_i se distribuye siguiendo una distribución conocida y definida para todo el dominio real. A su vez, se asume una independencia y hexogeneidad de v_i .

Los modelos pueden definirse como Ordered Logit o Probit de acuerdo a la distribución de ε_i asumida. En el modelo Logit se asume que la componente aleatoria sigue una distribución generalizada de valores extremos de tipo 1, también conocido como la distribución Gumbel, idénticamente e independientemente distribuida (GEV-1 IID). Mientras que en el caso del modelo Probit se asume que la variable se distribuye de acuerdo a la distribución Normal.

Centrándonos en el caso de aplicación de esta tesis, podemos definir la variable calidad del servicio como la variable latente a analizar q_i^* . Esta variable es evaluada por los usuarios escogiendo entre una de las opciones disponibles q_i :

- 0 Muy Mal
- 1 Mal
- 2 Normal
- 3 Bien
- 4 Muy bien

Cada evaluación realizada por los encuestados no proporciona el valor real de q_i^* , sino que nos muestra una versión restringida del mismo, una versión limitada a 5 opciones, siendo uno de ellos el más cercano al valor real percibido por el encuestado. Por lo tanto, el modelo describe las probabilidades de obtener los diferentes valores q_i . Además, se observa que no existe una relación directa entre las variables explicativas v_i y la variable dependiente q_i . Puesto que q_i es una mera etiqueta que se encuentra asociada a una franja de valores de probabilidad. La probabilidad asociada a la elección se define como:

$$\text{Prob}[q_i = j | v_i] = \text{Prob}[\varepsilon_i \leq \mu_j - \theta'v_i] - \text{Prob}[\mu_{j-1} - \theta'v_i], j = 0, 1, \dots, J \quad (18)$$

Para poder estimar los parámetros es necesario definir una serie de normalizaciones. En primer lugar, para mantener el signo positivo de todas las probabilidades es necesario que $\mu_j > \mu_{j-1}$. En segundo lugar, el modelo debe existir en todo el dominio real, por lo tanto, $\mu_{-1} = -\infty$ y $\mu_j = +\infty$. Por lo general los datos disponibles no muestran información sobre la escala de la variable dependiente (en cualquier caso, si el valor de escala de la variable y_i^* es modificada por un valor positivo cualquiera, entonces, modificando las escalas de las variables desconocidas μ_j y θ por el mismo valor las propiedades de las observaciones seguirían siendo iguales), no es posible estimar la varianza de la variable aleatoria $Var[\varepsilon_i] = \sigma_\varepsilon^2$. Para ello es recomendable establecer una restricción basado en $\sigma_\varepsilon = cte$. Es habitual definir el valor de la varianza como 1 para los modelos Probit y como $\pi^2/3$ para los modelos Logit. Finalmente, considerando la existencia de un término constante en el modelo es necesario establecer $\mu_0 = 0$.

El cálculo de los parámetros se realiza aplicando método de máxima verosimilitud (Pratt, 1981), siendo la función de log verosimilitud:

$$\log L = \sum_{i=0}^n \sum_{j=0}^J m_{ij} \log [F(\mu_j - \theta' \mathbf{v}_i) - F(\mu_{j-1} - \theta' \mathbf{v}_i)] \quad (19)$$

Donde $m_{ij} = 1$ si $q_i = j$ y 0 el resto de los casos y $F(\cdot)$ denota la función de distribución acumulada.

Es posible estimar modelos de datos ordenados considerando la heterogeneidad entre los distintos usuarios de forma similar a lo detallada en el apartado 3.1.2.

3.2.1 Modelos de datos ordenados con parámetros aleatorios

Una forma de considerar de incluir la aleatoriedad de los parámetros dentro de un modelo Ordered es considerar un modelo con heteroscedasticidad. El modelo se define de la siguiente forma:

$$\theta_i = \theta + \xi_i \quad (20)$$

Donde $\xi_i \sim N[\mathbf{0}, \Omega]$. Incluyendo esta expresión en la regresión de la variable latente, obtenemos:

$$q_i^* = \theta' \mathbf{v}_i + \varepsilon_i = \theta' \mathbf{v}_i + \varepsilon_i + \xi_i' \mathbf{v}_i \quad (21)$$

El resultado es un modelo Ordered Probit donde la perturbación tiene una varianza $Var[\varepsilon_i + \mathbf{v}_i' \xi_i] = 1 + \mathbf{v}_i' \Omega \mathbf{v}_i$. El modelo resultante obtiene.

$$\text{Prob}[q_i \leq j | \mathbf{v}_i] = \text{Prob}[\varepsilon_i + \xi_i' \mathbf{v}_i \leq \mu_j - \theta' \mathbf{v}_i] = F\left(\frac{\mu_j - \theta' \mathbf{v}_i}{\sqrt{1 + \mathbf{v}_i' \Omega \mathbf{v}_i}}\right) \quad (22)$$

En Boes and Winkelmann (2006) se sugiere que la estimación puede realizarse mediante medias ordinarias, aunque tiene que considerarse una nueva fuente de no-linealidad y por lo tanto, los elementos que componen Ω también deben de ser estimados. En William H Greene and Hensher (2010) se demuestra que es posible estimar los parámetros mediante la maximización de la verosimilitud simulada.

3.3 DISEÑO DE LA ENCUESTAS

3.3.1 Encuestas de Satisfacción

El objetivo de esta tesis es entender y analizar la calidad percibida por los usuarios de transporte público. Los datos necesarios para el análisis se consiguen mediante la realización de encuestas de satisfacción. Este tipo de encuestas se engloban dentro de las encuestas de preferencias reveladas (PR), cuya principal característica reside en que los encuestados son preguntados por una serie de características o variables que conocen y son capaces de valorar. Esto es, los usuarios responden preguntas sobre situaciones habituales o reales. En el caso concreto de las encuestas de satisfacción, los usuarios realizan una valoración del servicio de transporte público y de las distintas características que lo componen.

A la hora de diseñar una encuesta de satisfacción es importante considerar una serie de aspectos relevantes:

- Los atributos del servicio a evaluar deben ser definidos de forma que describan al máximo posible el servicio evaluado, sin extenderse de forma excesiva en el número de atributos definidos. Por ejemplo, un transporte público realizado mediante autobuses no será igual que un metro o un tranvía, y por lo tanto, los atributos que lo definan también deben ser distintos. Los atributos más representativos de un modo de transporte suelen definirse con estudios previos que aplican métodos de participación ciudadana como grupos focales o entrevistas en profundidad (Ibeas et al., 2011).
- La escala de medición de la satisfacción debe definirse para mantener una consonancia a lo largo de toda la encuesta. La escala puede ser cuantitativa (de 0 a 5, de 1 a 10, ...) o cualitativa (de “muy mal” a “muy bien”, de “totalmente en desacuerdo” a “totalmente de acuerdo”, ...). A su vez, será necesario definir el número de valores de la escala, esto es, el número de opciones de respuesta. Si la escala es muy amplia, digamos de 0 a 100, influye mucha complejidad para el encuestado a la hora de escoger un valor. Por el contrario, si la escala es muy simple, digamos de dos opciones “bien” y “mal”, la información obtenida sobre la satisfacción de los usuarios será muy pobre y no se podrá realizar un análisis apropiado. Lo habitual es definir encuestas de 5 puntos (escalas cualitativas y cuantitativas) o 10 puntos (más habitual en escalas cuantitativas).
- Por último, aunque el objetivo de las encuestas sea obtener la satisfacción de los usuarios, es recomendable incluir preguntas de carácter socioeconómico y del viaje para obtener información sobre el usuario encuestado. De esta forma es posible realizar análisis que relacionen los distintos tipos de usuarios con los distintos niveles de calidad percibida. Las variables más habituales son: sexo, edad, estado laboral, motivo del viaje, nivel de ingresos, disponibilidad de otro modo de transporte, etc. Al igual que para los atributos del sistema, las variables socioeconómicas dependerán de la realidad de la ciudad o región que se quiera analizar.

La obtención de datos suele realizarse a bordo de los sistemas de transporte, o en su defecto, en las paradas de los mismos. Como se ha mencionado en el análisis del estado del arte de esta tesis, los métodos de encuestado están siendo modificados para incluir nuevas tecnologías, sin embargo, la mayoría de los estudios realizados hasta la fecha utilizan el tipo de encuestado presencial. Para poder obtener unos resultados significativos de la realidad del servicio es necesario encuestar a un gran número de usuarios. Esta cuestión es respondida en el siguiente apartado donde se definen las fórmulas matemáticas más utilizadas para definir el tamaño de la muestra necesaria para obtener resultados representativos de la población global a analizar.

3.3.2 Encuestas Best-Worst

Las encuestas basadas en preguntas Best-Worst (o Best-Worst Scalling BWS) se basan en la idea de que un individuo debe enfrentarse a una elección entre alternativas compuestas por tres o más opciones, y es capaz de identificar la mejor y la peor opción de entre las alternativas (Louviere et al., 2015). Las palabras “Best” (mejor) y “Worst” (peor) son meramente representativas de las opciones a escoger y pueden ser sustituidas por cualquier término que defina los extremos de una variable latente continua. Por ejemplo, pueden definirse escalas de importancia definiendo las opciones “más importante” y “menos importante”.

Se definen un total de 3 casos distintos, los cuales solamente difieren en la complejidad de las alternativas mostradas en cada pregunta. Estos casos son: Caso 1 elección entre objetos (The object case), Caso 2 elección entre perfiles (The profile case) y el Caso 3 la elección multi-perfil (The multi-profile case). Estos distintos casos se desarrollan a continuación.

3.3.2.1 *Caso 1 La elección entre objetos*

Este primer caso se trata del modelo clásico desarrollado por Louviere a finales de los años 80 (Finn and Louviere, 1992). El objetivo de este cuestionario es medir la relación de una serie de objetos en base a una escala subyacente, latente y subjetiva. Estos objetos pueden tratarse de prácticamente cualquier cosa: distintos productos, atributos de un servicio, marcas comerciales, personas, etc. De la misma forma, las mediciones pueden representar escalas totalmente distintas, como puede ser: calidad de los objetos, importancia, preferencia de uso, disposición al pago, etc.

Por lo tanto, el caso 1 requiere de disponer de una lista de objetos o ítems que han de ser medidos acorde a una escala definida. La lista es exógena del ejercicio BW, pero los elementos que lo componen deben mostrar relación entre ellos.

Los elementos que componen la lista completa son agrupados en sub grupos que son los mostrados en cada pregunta. Estos subgrupos suelen estar compuestos por 4 o 5 objetos debido a que un mayor número de elementos puede dificultar al encuestado responder a la pregunta. Las preguntas son realizadas en una encuesta, donde el encuestado debe

seleccionar la opción “best” y la opción “worst” de entre todos los objetos del subgrupo de acuerdo a la escala que se plantea. En la Tabla 1 se puede ver un ejemplo de BW caso 1 donde el encuestado debe escoger de entre una serie de modos de transporte. Puesto que en el caso 1 no se asocia ningún tipo de atributo a los objetos, la elección de uno o de otro dependerá exclusivamente de las características del encuestado y su experiencia previa en elecciones similares.

Tabla 1 Ejemplo de pregunta BW Caso 1

Mejor modo de transporte (Escoger una opción ■)	Modos disponibles	Peor modo de transporte (Escoger una opción ■)
<input type="checkbox"/>	Coche	■
■	Autobús	<input type="checkbox"/>
<input type="checkbox"/>	Metro	<input type="checkbox"/>
<input type="checkbox"/>	Bicicleta	<input type="checkbox"/>

El principio subyacente de la elección BW reside en el mismo que en el de los experimentos de elección discreta, esto es, la teoría de la utilidad aleatoria. Esta teoría asume que los usuarios cometen errores a la hora de responder a las preguntas, sin embargo, al plantearles una serie de preguntas sucesivas similares es posible, debido a la frecuencia de respuesta, obtener una indicación del valor real de cada objeto evaluado. Además, combinando respuestas de varios individuos donde un objeto A es escogido más veces que un objeto B, es posible cuantificar la preferencia de A sobre B.

La necesidad de definir unos subgrupos reducidos de 4 o 5 objetos requiere establecer un criterio de diseño que permita obtener los mejores resultados. Teniendo en cuenta, que la lista de objetos completa se compone de un gran número de objetos, puede darse el caso que no sea posible mostrar todos los objetos a todos los encuestados. Por lo tanto, es importante definir el método de diseño de los subgrupos. Una forma de hacerlo es mediante la elección aleatoria de los objetos, escogiendo los objetos pertenecientes a cada subgrupo de forma completamente aleatoria. Este método puede utilizarse, pero es más recomendable para los casos donde el número total de objetos pueda ser abarcado por cada encuestado. En caso contrario, si el número de variables es demasiado grande, se recomienda realizar un diseño eficiente más avanzado (Huber and Zwerina, 1996; Rose and Bliemer, 2009).

3.3.2.2 *Caso 2 La elección entre perfiles*

La elección del caso 2 se refiere a los niveles de atributos definidos como perfiles. Un perfil se define como una combinación de los niveles de los atributos, más comúnmente puede asociarse a un único escenario de cálculo diseñado para una encuesta de Preferencias Declaradas (PD) habitual. Donde a cada atributo se le asocia uno de los valores que puede tener ese atributo. Dicho de otra forma, un objeto (modo de transporte, marca comercial, persona, etc.) puede describirse mediante una serie de atributos asociados al mismo (características, tamaño, precio, etc.). Cada atributo puede asociarse a uno o más valores llamados niveles, que suelen abarcar todo el rango del ámbito de

estudio. Por ejemplo, si consideramos el tiempo de viaje de un modo de transporte, es necesario que en la encuesta se considere desde el menor tiempo hasta el mayor tiempo posible para realizar el mismo trayecto considerando todas las opciones disponibles, sean realmente existentes o futuros desarrollos. Los niveles pueden ser cuantitativos y (más o menos) continuos, como el precio o el tiempo; o tener un carácter discreto, como el género. Cada atributo se define por un tipo y valores de nivel propios. Di bien, algunos objetos pueden compartir un mismo atributo, pero tener distintos niveles asociados. Por ejemplo, el tiempo de viaje es un atributo definido en todos los modos de transporte disponible, pero es muy posible que los niveles definidos para el coche sean distintos a los niveles definidos para la bicicleta, puesto que el coche privado es un vehículo considerablemente más rápido. Un ejemplo de una pregunta BW de caso 2 se representa en la siguiente tabla (Tabla 2).

Tabla 2 Ejemplo de pregunta BW Caso 2

Considero que este atributo es el más atractivo (Escoger una opción ■)	Atributos de los modos disponibles	Detalles del atributo	Considero que este atributo es el menos atractivo (Escoger una opción ■)
<input type="checkbox"/>	Tiempo de viaje	15 min	<input type="checkbox"/>
<input type="checkbox"/>	Precio	2 €	<input type="checkbox"/>
<input type="checkbox"/>	Número de paradas	2	<input type="checkbox"/>
<input type="checkbox"/>	Clima	Lluvia	■
■	Estado del tráfico	Poco tráfico	<input type="checkbox"/>

3.3.2.3 *Caso 3 La elección multi-perfil*

El tercer y último caso se puede asociar a un experimento de elección discreta típica. En cada escenario al encuestado se le muestran una serie de alternativas las cuales se componen de atributos con distintos niveles, esto es, distintos perfiles. En la Tabla 3 se muestra como para cada alternativa o modo de transporte elegible, los niveles de los atributos varían, siendo de esta forma más o menos atractivos para el encuestado.

En los experimentos de elección discreta tradicionales, los encuestados deben escoger solo la opción más atractiva de acuerdo a la descripción de los atributos que se muestran. En el caso de las preguntas BW no solo se obtiene la respuesta más favorable (“BEST”) sino que se obtiene aquella que se considera más desfavorable o menos atractiva (“Worst”), duplicando la información conseguida en cada pregunta.

Tabla 3 Ejemplo de pregunta BW Caso 3

Atributos	Coche	Autobús	Metro	Bicicleta
Tiempo de viaje	20 min	30 min	15min	40 min
Precio	0,6 €/h parking	0,9 €	1,5 €	-
Número de paradas	-	3	5	-
Clima	Lluvia	Soleado	Soleado	Lluvia
Estado del tráfico	Congestión	Poco tráfico	Normal	Normal
Opción más atractiva (Escoger una opción ■)	<input type="checkbox"/>	<input type="checkbox"/>	■	<input type="checkbox"/>

Opción **menos** atractiva
(Escoger una opción ■)



Como se ha mencionado anteriormente, los perfiles son definidos de acuerdo a unos atributos y a unos niveles. Estos perfiles componen un escenario de elección, el cual ha de ser diseñado para incluir todos los niveles de los distintos atributos a lo largo de los distintos escenarios propuestos en una encuesta. Si el número de atributos y los niveles asociados a los mismos es muy alto, no es posible mostrar todas las combinaciones posibles a cada encuestado. Por ello, es necesario diseñar los escenarios de forma adecuada para que la combinación de los niveles permita comparar los resultados de forma eficiente. Es habitual en este tipo de encuestas realizar el diseño mediante el diseño eficiente minimizando el D-error, método que se expone más detalladamente en el siguiente apartado.

34 TAMAÑO MUESTRAL

En teoría, si el objetivo es estudiar la satisfacción de los usuarios de un transporte público, sería necesario realizar una encuestada a cada uno de esos usuarios. Sin embargo, en la práctica, realizar encuestas a todos los usuarios no suele ser posible. Para ello, es necesario definir un tamaño de muestra, un grupo reducido de usuarios, que permita estimar la satisfacción de todos los usuarios con el menor error posible.

La definición del modo de cálculo para el tamaño de la muestra depende del método de encuestado utilizado. En el caso de las encuestas de satisfacción, el método habitual es de preferencias reveladas. Donde los encuestados deben responder unas preguntas que acotan su posible nivel de satisfacción a una escala de distintos valores. Puesto que las encuestas consideran distintas variables, el tamaño de la muestra debe estar definido por la variable más restrictiva.

Considerando que todos los usuarios de TP componen la población N , podemos asumir que una variable específica se distribuya con media μ y desviación estándar σ . En cuanto a la muestra encuestada n , la misma variable tendrá una media \bar{x} y una desviación estándar σ_x . El teorema central del límite postula que la estimación de la media de una muestra tiende a una distribución Normal siempre y cuando la muestra sea mayor o igual a 30. Por lo tanto, se define que la distribución de la media \bar{x} es una distribución normal de media μ y desviación estándar $se(\bar{x})$, también conocida como error estándar de la media. El error estándar puede calcularse de la siguiente forma:

$$se(\bar{x}) = \sqrt{\frac{(N-n)\sigma^2}{(N-1)n}} \quad (23)$$

Para tamaños de población muy grandes, el factor $\sqrt{\frac{(N-n)}{(N-1)}}$ es muy próximo a 1. Quedando la expresión anterior:

$$se(\bar{x}) = \frac{\sigma}{\sqrt{n}} \quad (24)$$

Si se considera solo la muestra, entonces, la mejor estimación de los parámetros μ y σ vienen dados por los parámetros \bar{x} y σ_x . La ecuación anterior puede definirse como:

$$se(\bar{x}) = \frac{\sigma_x}{\sqrt{n}} \quad (25)$$

Por lo tanto, el cálculo de la muestra necesaria solo depende de dos factores, la desviación estándar de la muestra y el error estándar asociado a la media. Aunque esto puede ser sencillo a simple vista, existen dos inconvenientes que dificultan su aplicación. El primero reside en la estimación de la varianza de la muestra σ_x , puesto que este solamente puede calcularse una vez se han obtenido los datos de la muestra, por lo que es necesario estimarlo mediante otras fuentes. El segundo problema está relacionado con el error estándar y al grado de confianza que se desea otorgar la media de la muestra \bar{x} como estimador de la media poblacional μ . La práctica habitual consiste en definir un intervalo de error alrededor de la media para un nivel de confianza definido. Por lo tanto, es necesario definir el nivel de confianza aceptable (comúnmente se utiliza un nivel de confianza del 95%, lo que implica que se comete un error el 5% de los casos) y los límites del intervalo alrededor de la media, ya sean en términos absolutos o relativos (es habitual considerar un error asociado del 10%).

Si se considera una distribución normal, los valores pertenecientes a un intervalo concreto pueden definirse como $\mu \pm Z_\alpha \sigma$. Para un nivel de confianza del 95% $Z_\alpha = 1,96$, lo que significa que el 95% de los valores se encontrarán dentro de la banda comprendida entre $\mu \pm 1,96\sigma$. Especificando un error del 10%, entonces los valores asumibles serán aquellos comprendidos en $\mu \pm 0,1\mu$. Asumiendo que la desviación estándar de la media corresponde con el error estándar, entonces, $Z_\alpha \cdot se(\bar{x}) = e \cdot \mu$, siendo e el error asumido, que con la media conforman el parámetro de precisión que se define como d . Combinado esto con la expresión anterior conseguimos que el tamaño de muestra mínimo quede definido de la siguiente forma.

$$n = \frac{Z_\alpha^2 \sigma_x^2}{d^2} \quad (26)$$

O en el caso de tratarse de una población finita.

$$n = \frac{NZ_\alpha^2 \sigma_x^2}{d^2 (N-1) + Z_\alpha^2 \sigma_x^2} \quad (27)$$

Sin embargo, en la práctica, la falta de datos no permite disponer de la desviación estándar de la muestra con antelación a la realización de la encuesta. Por ello es habitual utilizar una fórmula sustitutiva que permite calcular el número de encuestas requeridas sin necesidad de suponer ningún valor. Para ello se asume lo siguiente, de las n encuestas realizadas, X_1 escogen una puntuación concreta. X_1 tiene una distribución hipergeométrica $H(N, N', n)$ donde N' representa el número total de viajeros que escogen la puntuación concreta. La esperanza matemática y varianza se definen como:

$$\begin{aligned} E(X_1) &= np \text{ con } p = N' / N \\ \text{Var}(X_1) &= np(1-p)(1-n/N) \end{aligned} \quad (28)$$

Aplicando el teorema central del límite:

$$X_1 \sim N(np, np(1-p)(1-n/N)) \quad (29)$$

Es posible definir el estimador de p como:

$$\hat{p} = \frac{X_1}{n} \quad (30)$$

En consecuencia:

$$\hat{p} \sim N\left(p, \frac{p(1-p)(1-n/N)}{n}\right) \quad (31)$$

Como se ha comentado anteriormente el intervalo de confianza viene definido por el valor Z_α y cuyo valor no debe superar el error asumido. Esto quiere decir que:

$$Z_\alpha \sqrt{\frac{p(1-p)(1-n/N)}{n}} \leq e \quad (32)$$

Lo que equivale a:

$$n \geq \left(\frac{e}{Z_\alpha} \right)^2 \frac{p(1-p)}{p(1-p)} + \frac{p(1-p)}{N} \quad (33)$$

Debido a que el valor de p es desconocido la práctica habitual es definir los valores más conservadores para un error del 10% y un nivel de confianza del 95%. Esto es, un valor de $p=0,5$.

En el caso de las encuestas PD, el tamaño de la muestra viene definido por el método de diseño empleado. Siendo el más comúnmente utilizado el diseño eficiente expuesto en el apartado anterior.

35 MÉTODOS UTILIZADOS EN CADA ARTÍCULO

3.5.1 Artículo 1

El objetivo principal del primer artículo presentado en el compendio fue el de analizar la influencia de la importancia de los atributos de un sistema de transporte público a la hora de explicar la satisfacción general del mismo. Para ello, se propuso un estudio realizado en dos fases. Una primera fase de recolección de datos y una segunda de modelización.

Para la recolección de datos, se planteó una encuesta de satisfacción que se realizó a bordo de los autobuses del servicio de transporte público de la ciudad de Santander. La encuesta se dividió en 3 partes. En la primera fase, los usuarios encuestados debían responder a una serie de preguntas socioeconómicas escogiendo una de las opciones de respuesta disponibles. Las variables socioeconómicas requeridas se especifican en la Tabla 4.

Tabla 4 Variables socioeconómicas obtenidas en la encuesta de satisfacción de 2015 (Artículo 1)

Género
Edad
Situación laboral
Carnet de conducir
Posesión de vehículo propio
Motivo del viaje
Frecuencia de uso (Viajes realizados por semana)
Método de pago
Salario mensual

La segunda parte hacía referencia a los distintos atributos que componen el servicio de transporte público. En este caso, los encuestados debían valorar cada atributo siguiendo una escala Likert de 5 puntos (Muy mal, Mal, Normal, Bien o Muy Bien). Consiguiendo así la satisfacción de los usuarios con los distintos aspectos del servicio. Los distintos atributos evaluados se muestran en la Tabla 5, los cuales fueron agrupados en 6 grupos.

Tabla 5 Atributos del Servicio evaluados en la encuesta de satisfacción de 2015 (Artículo 1)

Nivel de servicio	Tiempo de acceso a la parada
	Tiempo de espera
	Tiempo de viaje
	Tiempo desde la parada al destino final
	Precio del billete
Servicio ofertado	Facilidad de transbordo
	Servicio ofertado (horarios, frecuencias)
	Fiabilidad del Servicio
	Líneas especiales para eventos
	Servicios nocturnos / fin de semana
	Cobertura de las líneas
Información	Información en las paradas
	Información en las plataformas digitales
	Información a bordo
Confort	Nivel de ocupación
	Aire acondicionado / calefacción
	Espacio para personas con movilidad reducida
	Confort de los autobuses
	Limpieza de los autobuses
	Posibilidad de portar grandes objetos: tablas de surf, maletas, animales, etc.

Comportamiento del conductor	Forma de conducción
	Amabilidad del conductor
Sostenibilidad	Uso de autobuses híbridos
	Ruido

La evaluación de los atributos se realizó por etapas, mostrando de forma simultanea solo aquellos atributos pertenecientes a un mismo grupo. Una vez realizada la evaluación, se pedía al encuestado que ordenase los atributos pertenecientes a un mismo grupo de mayor a menor importancia.

En la parte final de la encuesta, los encuestados debían evaluar el servicio de transporte público en su conjunto siguiendo la misma escala Likert empleada con los atributos. Además, se mostraban los nombres de los 6 grupos y los encuestados debían ordenarlos de mayor a menos importancia.

Con los datos obtenidos se procedió a la modelización. Se realizaron un total de 6 modelos Ordered Probit. El primer modelo se trata de un modelo sencillo que relaciona de forma directa la satisfacción de los atributos con la satisfacción general del servicio. El segundo modelo mejoraba el primero incluyendo las variables socioeconómicas de los encuestados. De esta forma se capturaba parte de la heterogeneidad de los usuarios en el modelo. El tercer modelo, además de incluir la heterogeneidad de la muestra, añadía la importancia explícita obtenida en la encuesta, combinando la importancia relativa de cada grupo con la importancia relativa de los atributos pertenecientes al mismo. Por último, para modelizar la heterogeneidad de la muestra de forma más avanzada, se replicaron estos tres modelos, pero considerando la posibilidad de estimar parámetros aleatorios.

3.5.2 Artículo 2

El segundo artículo perteneciente al compendio busca demostrar que es posible obtener los mismos resultados de un análisis de satisfacción en un sistema de transporte público partiendo de una base de datos menor. Esto es, reduciendo la cantidad de preguntas realizadas en cada encuesta. Para ello se utilizó la base de datos obtenida en la encuesta realizada en el artículo 1.

Partiendo de la base de datos completa, se eliminó la mitad de la información disponible en cada encuesta de forma aleatoria. Más concretamente, se eliminaron 12 de las 24 evaluaciones realizadas a los atributos del sistema. Para cada observación, la selección de los 12 atributos eliminados fue aleatoria. Simulando de esta forma, una encuesta ficticia en que se hubieran mostrado solo 12 atributos a cada encuestado, escogidos estos de forma aleatoria de los 24 totales. La parte de caracterización y la valoración general del servicio se mantuvieron iguales a la base de datos completa.

El análisis de la viabilidad de esta encuesta reducida para proporcionar resultados similares a la encuesta completa se realizó en dos fases. En una primera fase se compararon los resultados estadísticos básicos de la muestra reducida y la muestra

completa. Dicho de otra forma, se estudió que las medias, las modas y las desviaciones de las valoraciones de los atributos fueran similares para las dos bases de datos.

En la segunda fase se estudiaron los resultados de la modelización. El método de modelización fue el mismo que se aplicó en el estudio previo mostrado en el artículo 1, el modelo Ordered Probit. En este estudio en particular se utilizó el modelo sencillo sin incluir la variación sistemática o aleatoria de la muestra. Sin embargo, una de las necesidades del modelo es que para su estimación es necesario disponer de una base de datos completa de todas las observaciones. Esto es, que todos los encuestados hayan evaluado el total de los atributos incluidos en el modelo.

Debido a esta necesidad del proceso de estimación, fue necesario completar la base de datos reducida mediante imputación. En el estudio se consideraron distintos métodos de imputación con el objetivo de comparar los resultados y escoger aquel método que mejor resultado ofrezca en comparación al modelo estimado con la base de datos completa.

Se estudiaron tres métodos de imputación. El primer método consistió en completar la base de datos mediante el uso de la moda de todas las evaluaciones de cada atributo, dicho de otra forma, la respuesta más común para cada atributo de acuerdo a las respuestas disponibles. El segundo método de imputación se basó en estimar un modelo Ordered Probit por cada atributo. Donde las variables independientes son las características socioeconómicas de cada encuestado y la variable dependiente es cada uno de los 24 atributos. El tercer y último método consistió en aplicar la imputación múltiple basado en Fully Conditional Specification (FCS), que utiliza un método iterativo Monte Carlo con cadenas de Markov.

3.5.3 Artículo 3

El objetivo de la tercera investigación presentada era estudiar la importancia que tienen distintos atributos del ferrocarril. Probando en el proceso el tipo de encuesta Best-Worst caso 1. El proceso metodológico se dividió en dos fases diferenciadas. Una fase cualitativa y una fase cuantitativa.

La fase cualitativa consistió principalmente en obtener información sobre el estado actual de ferrocarril y los aspectos más importantes del mismo. El estudio cualitativo se realizó mediante un proceso consultivo a diferentes agentes involucrados en el ámbito ferroviario. Para ello se realizaron varios grupos focales con usuarios del ferrocarril y trabajadores. La información cualitativa se completó con entrevistas en profundidad a expertos ferroviarios.

Mediante la información cualitativa se definieron un total de 30 variables que sirven para definir el ferrocarril. Estas variables (Tabla 6) fueron posteriormente utilizadas en la fase cuantitativa.

Tabla 6 Atributos incluidos en la encuesta de importancia ferroviaria (Artículo 3)

Código	Atributos
1	Información para acceder a la estación

2	Tiempo de acceso a la estación
3	Tiempo de transbordo de un tren a otro
4	Tiempo de viaje en tren (duración del viaje)
5	Confort a bordo (asientos cómodos, ergonomía y servicios a bordo)
6	Disponibilidad de aparcamiento para coches en la estación
7	Disponibilidad de aparcamiento para bicicletas en la estación
8	Información sobre conexiones con el transporte urbano en la estación
9	Información sobre conexiones con el transporte interurbano en la estación
10	Número de trenes por día u hora (frecuencia)
11	Servicios a bordo (restaurante, coches cama, guardería, vagones silencio, Wi-Fi, enchufes, etc.)
12	Información a bordo (conexiones y transbordos, paradas, incidencias, etc.)
13	Política ambiental de la compañía
14	Tarifas (política tarifaria, descuentos, servicios low-cost, etc.)
15	Segregación a bordo (1er y 2da clase)
16	Servicios en el idioma local y en inglés (información y servicios disponibles en varios idiomas)
17	Sistema de billeteaje (billete electrónico, billetes integrados con otros medios de transporte, etc.)
18	Seguridad (a bordo frente a accidentes)
19	Seguridad (Controles policiales en las estaciones, control de equipaje, etc.)
20	Posibilidad de transportar bicicletas a bordo
21	Posibilidad de transportar animales a bordo
22	Posibilidad de transportar equipaje sin límite de peso
23	Accesibilidad para personas de movilidad reducida
24	Posibilidad de reservar asiento
25	Facilidad de compra del billete (más puntos de venta, máquinas de venta de billetes, compra online, etc.)
26	Duración de los controles de seguridad para acceder al tren (control de equipaje)
27	Conexiones con el transporte urbano en la estación (para acceder y para dejar la estación)
28	Conexiones con el transporte interurbano en la estación (para acceder y para dejar la estación)
29	Localización de las estaciones en las ciudades
30	Información sobre actividades en el destino (economía, ocio, servicios en el destino, etc.)

La fase cualitativa de esta investigación consistió en la realización de una encuesta de tipo Best-Worst Caso 1, donde los encuestados debían seleccionar el atributo más importante de entre un conjunto de 4 alternativas. La encuesta se realizó utilizando una plataforma de encuestado online y se difundió mediante mailing y redes sociales. En la Figura 2 se muestra un ejemplo de una pregunta utilizada en la encuesta.

Scenario 3

From the following railway characteristics, choose the one you consider is the most important and the one you consider least important.

	Most Important	Least Important
Car parking availability at the station	<input type="radio"/>	<input type="radio"/>
Travel time to get to the station	<input type="radio"/>	<input type="radio"/>
Fares (Fare policy, discounts, low cost services...)	<input type="radio"/>	<input type="radio"/>
Number of daily and hourly trains	<input type="radio"/>	<input type="radio"/>

Figura 2 Ejemplo de pregunta Best-Worst de la encuesta de importancia ferroviaria (Artículo 3)

Con los datos obtenidos en la encuesta se estimó un modelo Logit multinomial basado en respuestas BW. De esta forma se consiguió establecer un orden de importancia cuantitativo a las 30 variables definidas en la fase cualitativa.

3.5.4 Artículo 4

El objetivo de la cuarta investigación presentada en esta tesis es la de comparar los dos métodos de encuestados tratados en los tres artículos anteriores. Por una parte, se plantea la estimación de un modelo Ordered Logit basado en valoraciones cualitativas. Y por otra, una modelización mediante unos modelos Logit Multinomial basado en respuestas Best-Worst.

Para realizar los dos análisis se diseñó una nueva encuesta que incluída los dos tipos de respuesta para cada atributo evaluado. En primer lugar, se actualizó la lista de atributos utilizada en estudios anteriores definiendo un total de 24 atributos (Tabla 7).

Tabla 7 Atributos de Servicio incluidos en la encuesta de satisfacción de 2017 (Artículo 4)

Orden	Atributo
1	Autobuses híbridos
2	Tiempo de acceso a la parada
3	Tiempo desde la para al destino final
4	Limpieza
5	Facilidad de transbordo
6	Información en las paradas
7	Información a bordo

8	Confort de los buses
9	Fiabilidad del servicio
10	Amabilidad del conductor
11	Calidad de las paradas
12	Información en aplicación móvil
13	Cobertura de las líneas
14	Información en la página web
15	Espacio para personas de movilidad reducida
16	Tiempo de espera
17	Nivel de ocupación
18	Claridad del mapa de servicios
19	Tiempo de viaje
20	Frecuencia de los servicios y horarios
21	Forma de conducción
22	Precio del billete
23	Calefacción / aire acondicionado
24	Ruido

A cada encuestado se le mostraban la mitad de estos atributos en bloques de 4 atributos que se repetían 3 veces. En ningún caso se repetía un mismo atributo para un mismo encuestado. La elección de los 12 atributos se realizaba de forma aleatoria balanceando el número de veces que cada atributo era mostrado a un usuario para obtener un número de observaciones homogénea para cada atributo.

En cada bloque, los encuestados debían realizar dos acciones, por una parte, debían evaluar los atributos mostrados siguiendo una escala cualitativa de 5 puntos. Por otra, considerando el mismo bloque de atributos mostrados, los encuestados debían escoger el atributo con el que más y menos satisfechos se encontraban y aquel que consideraban de mayor y menor importancia. En la siguiente figura (Figura 3) se muestra un ejemplo de un bloque tipo. Este bloque se repetía 3 veces para cada usuario. Al finalizar la encuesta, el encuestado debía evaluar el servicio en su conjunto mediante la misma escala cualitativa.

Example	Rating scale (5 point likert scale)						Best-Worst choice			
	Very Bad	Bad	Normal	Good	Very Good	N/A	Most Important	Least Important	Most Satisfied	Least Satisfied
Attribute 1					X		X		X	
Attribute 2			X							
Attribute 3				X				X		
Attribute 4			X							X

Figura 3 Ejercicio de respuesta de la encuesta de satisfacción de 2017 (artículo 4)

Una vez obtenidos los datos se procedió a la modelización. El modelo Ordered Logit se realizó siguiendo el mismo método propuesto en el artículo 2 de este mismo compendio, puesto que la base de datos disponible no se encontraba completa. A su vez, se estimaron dos modelos separados de Logit multinomial (MNL), uno para la respuesta BW de

satisfacción y otra para la respuesta de importancia. El método de modelización empleado para los modelos basados en BW fue el mismo que el empleado en el artículo 3 de este compendio. Adicionalmente, se realizó un análisis estadístico básico de los resultados de la evaluación obteniendo las medias, modas y desviaciones de las satisfacciones de todos los atributos.

Una vez obtenidos todos los resultados se realizó la comparativa. En primer lugar, era necesario definir el significado de cada uno de los modelos para poder compararlos acorde con su significado. Por una parte, se disponía de información relativa a la satisfacción de los usuarios referente a los distintos atributos, siendo estos, los resultados del análisis estadístico básico y el modelo MNL basado en las respuestas BW de satisfacción. Por otra parte, se disponía de información referente a la importancia de los atributos, basado en los resultados del modelo Ordered Logit y los resultados del modelo MNL basado en las respuestas BW de importancia. De forma que, se compararon los valores normalizados de los resultados para observar si la escala de los atributos y su variabilidad era similar en ambos métodos. Tanto para el caso de la satisfacción (comparando el análisis estadístico frente a los valores de los parámetros estimados del modelo MNL para los datos BW de satisfacción), como para el caso de la importancia (comprando los parámetros estimas con el modelo Ordered Logit con los parámetros estimados del modelo MNL con datos BW de importancia).

Por último, se planteó la posibilidad de poder obtener los resultados de un método utilizando los resultados obtenidos mediante el otro. Más concretamente, se planteó la posibilidad de obtener los resultados del estudio estadístico de la satisfacción partiendo de los datos BW. Al igual que la de obtener los valores de los parámetros del modelo Ordered Logit partiendo de los parámetros del modelo MNL estimado con datos BW. De forma resumida, se estudió la posibilidad de conseguir los resultados del método más común de encuestado partiendo de una encuesta de tipo BW. Para ello, se estimaron sendos modelos de regresión lineal.

“Algunas sesiones son estrellas y otras son piedras, pero al final todas son rocas y construimos sobre ellas.”

“Some sessions are stars and some sessions are stones, but in the end they are all rocks and we build upon them.”

(Wellington and Aylwin, 2013)

CAPÍTULO 4: COMPENDIO DE ARTÍCULOS

4.1 ARTÍCULO 1: MODELLING PERCEIVED QUALITY FOR URBAN PUBLIC TRANSPORT SYSTEMS USING WEIGHTED VARIABLES AND RANDOM PARAMETERS

4.1.1 Indicadores bibliométricos de la revista de publicación del artículo 1 en el JCR



Título de la revista: Transport Policy

ISSN: 0967-070X

eISSN: 1879-310X

Editorial: Elsevier SCI LTD

País de publicación: Estados Unidos

Volumen de publicación: 67

Fecha de publicación: 15 de septiembre de 2018

Páginas: 31-39

DOI: <https://doi.org/10.1016/j.tranpol.2017.05.006>

Factor de impacto JCR

Tabla 8 Factores de impacto de la revista del artículo 1

Año de publicación (2018) ¹	3,190
Promedio de 5 años	3,430

Ranking

Tabla 9 Ranking de la revista del artículo 1

Área de conocimiento	Posición	Percentil	Cuartil
Economics	45/363	87,741	Q1
Transportation	11/36	70,833	Q2

Numero de citas que posee el artículo

Tabla 10 Número de citas del artículo 1

Plataforma de medición	Número de citas
Google Scholar	8
Web of Science	3
Scopus	7

¹ Se considerará como el valor del factor de impacto el establecido para el año de publicación del artículo. En el caso que el valor del año correspondiente no esté disponible se considerará el último año disponible.

4.1.2 Transcripción del artículo 1: Modelling perceived quality for urban public transport systems using weighted variables and random parameters

MODELING PERCEIVED QUALITY FOR URBAN PUBLIC TRANSPORT SYSTEMS USING WEIGHTED VARIABLES AND RANDOM PARAMETERS

Eneko Echaniz Beneitez

University of Cantabria, Spain

Luigi dell'Olio

University of Cantabria, Spain

Ángel Ibeas Portilla

University of Cantabria, Spain

ABSTRACT

In this paper a methodology is proposed to consider the importance attributed by users to individual variables in a perceived quality analysis. A two stage ranking based attribute survey is proposed. Firstly, the attributes belonging to each group are ranked and secondly, each group is ranked according to its importance.

A series of successive ordered probit models is proposed which also includes models considering systematic and random variations in user taste. The variables are weighted according to the individual and group rankings.

The article concludes that increasing the complexity of the models improves their capacity to represent reality, however, there comes a point when the effort required to obtain sufficient data to feed the complexity of the models is not efficient and the time taken is not compensated by the improved predictions.

Keywords

Perceived Quality, Ordered Probit, Urban Public Transport, Revealed Preference Survey.

1. INTRODUCTION

The analysis of user perceived quality for a public transport system is useful for defining possible improvements and maximising their influence on service quality. This article introduces a methodology based on econometric modelling which applies ordered probit models to establish the pertinent variables or attributes which have the greatest effect on overall service quality. Furthermore, various models have been analysed considering user heterogeneity and the importance of the variables.

The validity and utility of the proposed methodology have been tested in a real application: the public transport system in the city of Santander (Northern Spain), where the perceived quality has been determined for the 15 lines representing the city's public

bus service. Following the calculation of all the models, a model fit comparison was made to establish the benefits of using more complex models against simpler ones.

This article is divided into 6 sections. A brief state of the art review regarding public transport service quality is provided in section 2. The methodology followed is explained in section 3 which is further divided into three parts: the data collection process, the theoretical background of the models used in the study and an explanation of the models that were developed. The results from the practical application are presented in section 4 which is followed by a comparison between the estimated models. The article finishes with the main conclusions drawn from this research.

2. STATE OF THE ART

Service quality has been widely studied since it was first introduced by (Berry et al., 1990; Parasuraman et al., 1985), who defined the perceived quality of service as the difference between expected quality and the perceived service quality.

User perceived quality has been shown to have a positive effect on the demand for public transport services (Cascetta and Cartenì, 2014; Joewono and Kubota, 2007; Lai and Chen, 2011; Nilsson et al., 2001; Rojo et al., 2012; Tam et al., 2010; Wen et al., 2005). However, although users perceive a very good quality of service, this cannot be taken as a criteria of success of the public transport service and therefore cannot be used as the only reference when planning policies aimed at maintaining current usage and attracting new customers (de Oña et al., 2016b; Felleson and Friman, 2009).

The first step for studying the quality of a public transport service consists of finding its defining variables or attributes. (Parasuraman et al., 1988) defined the SERVQUAL scale composed of 22 valid attributes for evaluating quality in different services. Later, (Hensher, 2003) developed the SQI (Service Quality Indicators) scale, aimed at public passenger transport services, formed of 13 attributes. In similar research, the QUATTRO project (EC, 1999) developed a three level classification based on 8 main groups with a total of 99 attributes to be evaluated. Various authors have adapted these foundations for many specific case studies about the implications different attributes have for a service (de Oña et al., 2012; del Castillo and Benitez, 2012; Eboli and Mazzula, 2009; Eboli and Mazzulla, 2012; Kim and Chung, 2016; Metri, 2006).

The perception of quality in public transport services varies between the different kinds of users and variation in their tastes has been evaluated by using a diversity of methodologies. User characterisation for satisfaction surveys allows the researcher to associate the perception of quality to a specific type of user, classified according to their socioeconomic characteristics or those characteristics specifically related to the journey being made (de Oña et al., 2016d, 2016a; Diana, 2012; Filipović et al., 2009; Joewono and Kubota, 2007; Anil Minhans et al., 2015; Susilo and Cats, 2014). Another method for defining different user idiosyncrasies consists of considering that the variation in user perception can follow a statistical distribution which means that random parameters can be used to include it in modelling processes (Bordagaray et al., 2014; Hensher et al., 2010).

Other studies have tried to establish the importance users place on each service attribute (Baltes, 2003; Bolton and Drew, 1991; Cook and Kress, 1988; Garrido and Ortúzar, 1994; Grujičić et al., 2014; Nathanail, 2008). Although (Grujičić et al., 2014) used a multicriteria analysis to establish a relationship between the importance of a variable and the value of perceived quality, most research has not found a clear relationship between importance and the values given to attributes.

The international literature provides many examples of methodologies for modelling perceived quality. A methodology based on structural equations was developed in (de Oña et al., 2016d, 2013, 2015; Eiró and Martínez, 2014; Rahman et al., 2016). Another possibility was the application of decision trees (de Oña et al., 2012), which could be combined with an analysis of systematic variations through the use of clusters to generate models able to differentiate between different kinds of users (de Oña et al. 2016). Other methods for analysing perceived quality were not based on modelling but have provided some very interesting results, examples include descriptive statistics (Eboli and Mazzulla, 2011, 2007), multicriteria analysis (Nathanail, 2008), factor analysis (Fellsson and Friman, 2008) or neural networks (Garrido et al., 2014).

The methodology based on Ordered Probit models chosen for this research has proven to be an extremely versatile, efficient and useful tool for modelling perceived quality (Bordagaray et al., 2014; Çelik and Senger, 2016; dell'Olio et al., 2010; Hensher et al., 2010; Joewono and Kubota, 2007; Tyrinopoulos and Antoniou, 2008). This particular methodology allows ordered qualitative responses to be modelled, which means that the non linearity existing between the different replies can be considered (dell'Olio et al., 2010). Furthermore, this type of model is able to use interactions to incorporate systematic variations resulting from the different socioeconomic characteristics of users (Bordagaray et al., 2014). By considering randomness in model threshold parameters (Hensher et al., 2010) or in the parameters affecting the quality variables (Bordagaray et al., 2014), it becomes possible to consider the different socioeconomic characteristics and idiosyncrasies of the users assuming that they follow a known statistical distribution. However, no evidence has been found of the combination between systematic and random variations in the same model, nor of the inclusion of attribute importance within the model. Therefore, the present research aims to address this void and complete our knowledge about modelling perceived quality using Ordered Probit models which consider all the available information.

3. METHODOLOGY

3.1. Survey design

The representative quality variables for a public transport service have been defined from an analysis of the existing international bibliography and a series of focus groups (Ibeas et al., 2011) involving public transport users in the city of Santander. The resulting 24 variables are presented in the following table (Tabla 11). These variables have been grouped into 6 different clusters.

Tabla 11 Artículo 1: Table 1 – Public Transport Service attributes

Level of Service	Walking time to bus stop
	Waiting Time
Mejora en la recolección de datos y la modelización de la calidad percibida en sistemas de transporte público	

	Travel Time Time from the stop to final destination Ticket price
Supply	Ease of transfer Offered Service (Timetable, frequencies) Service reliability Special Lines for events, football, concerts, etc. Nocturnal / Weekend Services Coverage of lines
Information	Information at the bus stops Information at digital platforms Information on board buses
Comfort	Occupancy Air conditioning / heating system Space for people with reduced mobility Comfort of the buses (seats and grab handles) Bus cleanliness Possibility of carrying large objects: surfboards, luggage, animals, etc.
Driver behaviour	Driving style Driver kindness
Sustainability	Hybrid / Biofuel bus use Noise pollution

In addition, each one of the interviewees was asked to provide some characterisation information, as listed in the following table (Tabla 12). The possible replies were limited to a specific number of options, varying from one attribute to another.

Tabla 12 Artículo 1: Table 2 – Characterisation Variables

Gender
Age
Employment Status
Driving license
Car ownership
Trip Purpose
Number of trips made per week
Usual payment method
Monthly income

3.2. Theoretical Background

3.2.1. Ordered Models

The book *Modelling Ordered Choices: A Primer* (William H Greene and Hensher, 2010) was taken as a reference work for this section.

The latest form of the Ordered Probit model, based on regression, was proposed by McKelvey and Zavoina (1975, 1971) for the analysis of ordered, categorised and non quantitative choices.

Ordered models are based on dividing a continuous utility space in discrete bands using a threshold based system.

$$\begin{aligned}
 q_i^* &= \theta' v_i + \varepsilon_i, \quad i = 1, \dots, n, \\
 q_i &= 1 \text{ si } \mu_0 < q_i^* \leq \mu_1 \\
 &= 2 \text{ si } \mu_1 < q_i^* \leq \mu_2 \\
 &= 3 \text{ si } \mu_2 < q_i^* \leq \mu_3 \\
 &= \dots \\
 &= J \text{ si } \mu_{J-1} < q_i^* \leq \mu_J
 \end{aligned} \tag{34}$$

In a first approach, the assumption of constant coefficients and threshold parameters for all users is made. The key idea of the model resides in that the observations made are not just an accumulation of a number of discrete observation that might be ordered somehow, but the transformation of an only continuous variable that has to be ordered.

The models contains the marginal utilities θ , plus, $J+2$ threshold parameters, μ_j , all of them to be estimated using n observations. The collected data consist on the v_i independent variables of each observation and the results q_i . The model is completed by the random error parameter ε_i . It is assumed that the error ε_i is distributed according to a known CDF (Cumulative Distribution Function) distribution and defined in the whole real domain. This assumption also includes the independence and hexogeneity from v_i .

Focusing in the problem of this research. Let's suppose a band of five possibilities for each evaluation users have to do, where the options are:

- 0 Very Bad
- 1 Bad
- 2 Normal
- 3 Good
- 4 Very Good

Each evaluation q_i done by the user does not provide the real value of the evaluation q_i^* but a restricted version of it, a version limited to 5 options, one of which is the nearest to the real and exact quality perceived by the user.

The probabilities of the observed answers are:

$$\text{Prob}[q_i = j | v_i] = \text{Prob}[\varepsilon_i \leq \mu_j - \theta' v_i] - \text{Prob}[\varepsilon_i \leq \mu_{j-1} - \theta' v_i], \quad j = 0, 1, \dots, J \tag{35}$$

The model describes the probabilities of obtaining the different values. There is no obvious relationship between the dependant variable q_i and the independent variables v_i , because q_i is just a label that represents a probability band.

A number of normalisations, as discussed below, need to be established for estimating the model parameters.

Firstly, $\mu_j > \mu_{j-1}$ in order to maintain positive signs for all the probabilities. Secondly, the space studied by the model should be the whole real space, so $\mu_{-1} = -\infty$ and $\mu_J = +\infty$. Given that the data usually does not provide information about the scale of the dependant variable (in which case if the scale of q_i^* is modified by a positive value,

modifying the scale of the unknown μ_j and θ values by the same value, all the properties of the observations remain the same), meaning it is not possible to estimate the free variance $Var[\varepsilon_i] = \sigma^2_\varepsilon$. It is advisable to use a restriction based on $\sigma_\varepsilon = cte, \bar{\sigma}$. A variance equal to one is typically used for Probit models and equal to $\pi^2/3$ for Logit models. Finally, assuming the existence of a constant term in the model, it needs to be established that $\mu_0 = 0$.

The parameters of the model are calculated by applying a maximum likelihood estimation problem, (Pratt, 1981) and (Greene, 2007, 2008). The log likelihood function is:

$$\log L = \sum_{i=0}^n \sum_{j=0}^J m_{ij} \log[F(\mu_j - \theta'v_i) - F(\mu_{j-1} - \theta'v_i)] \quad (36)$$

Where $m_{ij} = 1$ if $q_i = j$ and 0 if not. The maximisation is constrained by $\mu_{-1} = -\infty$, $\mu_0 = 0$, and $\mu_j = +\infty$. The remaining constraint, $\mu_{j-1} < \mu_j$, can initially be imposed by using structural parameters.

$$\begin{aligned} \mu_j &= \mu_{j-1} + e^{\alpha_j} \\ &= \sum_{m=1}^j e^{\alpha_m} \end{aligned} \quad (37)$$

However, this should usually not be necessary.

In practice, the model's parameters are estimated using specific software such as NLOGIT, which was the package chosen for this research.

3.2.2. Random parameter Ordered models

The random parameter model is defined as:

$$\theta_i = \theta + u_i \quad (38)$$

Where $u_i \sim N(0, \Omega)$. Including this θ_i term in the regression model we get.

$$q_i^* = \theta_i'v_i + \varepsilon_i = \theta'v_i + \varepsilon_i + u_i'v_i \quad (39)$$

The result is an ordered probit model in which the disturbance has a variance of $Var[\varepsilon_i + v_i'u_i] = 1 + v_i'\Omega v_i$. This is an ordered probit model with heteroscedasticity.

Probability in the model is established as:

$$Pr ob[q_i \leq j | v_i] = Pr ob[\varepsilon_i + u_i'v_i \leq \mu_j - \theta'v_i] = F\left(\frac{\mu_j - \theta'v_i}{\sqrt{1 + v_i'\Omega v_i}}\right) \quad (40)$$

3.3. Models developed for this study

In this study 6 models have been developed in order to analyse how the complexity of the model affects its predictive capability.

The first model, or Base Model (BM), is based on a linear and direct relationship between the partial evaluation of the service attributes v_{ik} and the overall evaluation of the transport service v_i , where the parameter θ_k associated with each variable v_{ik} establishes the weight of each variable in the dependent variable q_i . The model includes the constant θ_0 in accordance with the explanation provided in section 3.2.1. The model is completed with the error term ε_{ik} associated with the different idiosyncrasies of the users. The mathematical expression is represented as follows:

$$q_i = \theta_0 + \sum_{k=1}^N \theta_k \cdot v_{ik} + \varepsilon_{ik} \quad (41)$$

with:

$$k \in [1, 2, \dots, N]$$

where:

q_i : Overall evaluation of user i

θ_0 : Model Constant

N : Number of evaluated attributes

θ_k : Coefficient of variable k

v_{ik} : Score of each user i for each variable k

The second model can be seen as an evolution of the BM model, considering systematic variations in user tastes (SVM). This was achieved by creating $n-1$ dummy variables z_{ic} for each of the categories described in Tabla 12, with n as the number of reply options within each category. Systematic variation was included by creating interaction between the dummy variables z_{ic} and the independent variables v_{ik} . The mathematical expression of this model is:

$$q_i = \theta_0 + \sum_{k=1}^N \theta_k \cdot v_{ik} + \sum_{k=1}^N \theta_{ck} \cdot z_{ic} \cdot v_{ik} + \varepsilon_{ik} \quad (42)$$

with:

$$k \in [1, 2, \dots, N]$$

$$z_{ic} = 0 \text{ o } 1$$

where:

θ_{ck} : Coefficient of the variable k interacting with variable c

z_{ic} : Characterisation coefficient related to a user i and a characterisation variable c

The third model, in addition to the systematic variations, considers the weighting of the variables (SVMW) derived from the user ranking and was performed over two stages. Firstly, each user created their ranking from the variables belonging to a certain group. Secondly, a ranking was created from the groups of variables (see section 4.1). The attributes within a group share some common similarities (Access time, Waiting time, Travel time...), while different groups represent general aspects of the service i.e. (Level of service, Information...) (see section 3.1).

This method provides two weighting variables; δ_{nk} for the weighting within a group and δ_{mk} for the weighting of the groups.

The values of these variables are normalised so that their values fall between 0 and 1. The normalisation depends on the total number of variables, where the value of 1 is associated with the variable which the user considers most important and 0 on the least important, however, the intermediate variables will have a different value depending on the total

number of variables within the same group. If a group only contains two variables, the more important one will take a value of 1 and that of lesser importance will take 0, whereas if the group is made up of three variables that of greatest importance will take the value of 1, least will take 0 and the remaining variable will take 0.5. If there are four variables the values will be 1, 0.66, 0.33 and 0, and so on. The weighting of the groups is performed in the same way. The resulting expression is presented below:

$$\begin{aligned}
 q_i = & \theta_0 + \sum_{k=1}^N \theta_k \cdot v_{ik} + \sum_{k=1}^N \theta_{ck} \cdot z_{ic} \cdot v_{ik} + \sum_{k=1}^N \theta_{nk} \cdot \delta_{nk} \cdot v_{ik} \\
 & + \sum_{k=1}^N \theta_{nmk} \cdot \delta_{nk} \cdot \delta_{mk} \cdot v_{ik} + \sum_{k=1}^N \theta_{nck} \cdot \delta_{nk} \cdot z_{ic} \cdot v_{ik} \\
 & + \sum_{k=1}^N \theta_{nmck} \cdot \delta_{nk} \cdot \delta_{mk} \cdot z_{ic} \cdot v_{ik} + \varepsilon_{ik}
 \end{aligned} \tag{43}$$

with:

$$k \in [1, 2, \dots, N]$$

$$z_{ic} = 0 \text{ o } 1$$

$$\delta_{nk} \in [0, 1]$$

$$\delta_{mk} \in [0, 1]$$

Where:

θ_{nk} : Coefficient of the variable k weighted by δ_{nk}

δ_{nk} : Weighting coefficient for every n variable k inside a group m

δ_{mk} : Weighting coefficient for every group m related to a variable k

θ_{nmk} : Coefficient of the variable k weighted by δ_{nk} and δ_{mk}

θ_{nck} : Coefficient of the simply weighted variable k interacting with variable c

θ_{nmck} : Coefficient of the double weighted variable k interacting with variable c

The remaining three models are extensions of the previous three with the addition of considering systematic and random variation in user tastes, by using randomly distributed parameters. The acronyms used for these models are RBM, SRVM and SRVMW. For the sake of brevity, the mathematical expressions of these models will be omitted, as they are coincident with those presented previously.

4. PRACTICAL APPLICATION

The basic aim of this study is to know the satisfaction levels of users of a public transport service. An additional goal was to compare the prediction efficiency of the different models. A series of surveys were required for the estimation of the models described in section 3.3. The surveys were asked in the city of Santander, a medium sized coastal town in Northern Spain.

4.1. Data collection

The surveys were asked on a group of bus lines managed by the local operating company TUS. A total sample size of 747 observations was obtained from across the 15 different lines. The most interviews were held on the lines with the highest demand following the distribution presented in Tabla 13.

Tabla 13 Artículo 1: Table 3– Number of surveys

Line	Number of completed surveys
L1	94
L2	93
L3	84
L4	63
L5C1	40
L5C2	40
L6C1	38
L6C2	37
L7C1	40
L7C2	39
L11	29
L12	38
L13	39
L14	38
L16	35
Total	747

Expression (11) was applied to estimate the minimum number of surveys needed for each line. Previous studies showed this to be a useful expression for calculating interception survey sample sizes (dell'Olio et al., 2011, 2010).

$$n \geq \frac{p(1-p)}{\left(\frac{e}{1-p}\right)^2 + \frac{p(1-p)}{N}} \quad (44)$$

The most conservative values were taken, where: $P=0.5$; $e=10\%$; $z=1.69$ and N =number of users of each line at rush hour.

The surveys were made throughout the day to users both on board the buses and waiting at the stops. The questionnaires were related to the journey the passenger had just finished, the journey they were currently on, or the journey they were about to make. The surveys were made with the aid of tablets to make them more interactive and attractive for the users and to facilitate the creation of the database to be used later.

The survey was divided into three parts. Firstly, a characterisation questionnaire was asked to obtain the data described in Tabla 12. The users were then shown the service attributes for their evaluation, however only the attributes belonging to one of the groups were shown together on each screen (see Tabla 11). The users were asked to score each attribute following a Likert qualitative scale with 5 options (Very Bad, Bad, Normal, Good and Very Good). They were then asked to order the variables belonging to the same group from greater to lesser importance. This process was repeated for the 6 groups. Finally, the user was asked to evaluate the service as a whole by providing their Overall

Quality following the same Likert scale as before. Furthermore, in this final part of the survey the interviewees were asked to order the 6 groups of variables from greater to lesser importance.

4.2. Estimated models

The models estimated below are useful in explaining the attributes influencing the user perception of overall quality of service. A total of 6 perceived quality models were estimated, the complexity of each model was modified to evaluate the improvement in their predictive capabilities. All the models were based on the sample taken from the 15 bus lines serving the city of Santander with a total of 747 observations. The specification of the estimated models is described in greater detail in section 3.3.

As many of the high number of variables were found to be not statistically significant they were not included in the model, given that they had no bearing on the model's predictive capacity and they had no influence on the dependent variable q_i . The variables relating to the comfort of the bus and the air conditioning/heating system were combined into the Comfort variable; similarly the three systems for providing information (Internet, Bus and Stop) have been combined into one Information variable. In both cases the combination was performed using a simple average. It was decided to perform this grouping because of the small number of significant variables present in these groups and the existence of high levels of correlation between them. Finally, as none of the interviewees scored the Overall Quality of the service as Very Bad, the scale was reduced to 4 choice possibilities.

4.2.1. Base Model (BM)

Tabla 14 Artículo 1: Table 4– Base Model (BM)

BM			
F.Log-Likelihood		-497.064	
AIC/N		1.371	
Count R ²		0.724	
Variable	Coef.	Test t	
Constant	-4.307	-11.53	
Walking time to bus stop	0.248	4.01	
Waiting Time	0.136	2.25	
Travel Time	0.202	2.87	
Ticket prices	0.078	1.54	
Ease of transfer	0.263	4.21	
Offered Service	0.064	0.98	
Service reliability	0.178	2.52	
Nocturnal / Weekend Service	0.132	2.22	
Line Coverage	0.250	3.91	
Information *	0.270	2.86	
Comfort **	0.579	6.22	
Driving style	0.285	4.38	
Threshold parameters			
μ ₁	1.745	17.35	

μ_2

4.606 33.67

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system.

To test the consistency of the model, the parameters of the independent variables need to be checked for the correct sign and their sufficiently high significance level. As an increase in the value of a service attribute is assumed to result in an increase in the Overall Quality the signs of the parameters need to be positive and this condition is observed by all the parameters. Another important question is the sign of the constant, as mentioned in section 3.2.1, the evaluations divide the space into segments, where the first segment, corresponding to the worst score, is found between the values of $-\infty$ and 0, so if all the parameters are positive and all the possible replies are positive then the constant has to be negative to produce a negative value. Referring now to the level of significance, it can be seen that almost all the parameters are considered to be statistically significant (Test $t > 1.96$) except the Price of Tickets and the Offered Service, however both variables have been kept because of their involvement in the management policies of the service.

Given that all the variables v_i are expressed in the same unit of measurement as the dependent variable q_i , the values of the parameters can easily be compared; this value represents the weight of each variable in the Overall Quality. The greater a parameter is the greater is its contribution to the latent quality variable. For this model Tabla 14 shows how comfort is the variable which has the greatest influence on Overall Quality, with double the value of the other parameters. However, it needs to be remembered that Comfort is an aggregate variable which includes the combined effect of a group of variables as does Information. The combined effect of a group of variables forming The Level of Service and the Supply show that a service with good accessibility, good journey times and reliability represents a very good quality service.

The fit of the model is determined using the Log likelihood coefficient, AIC/N and Count R^2 . This model has served as the starting point for analysing the improvement in the predictions of the other models. More information can be found in section 5 where a comparative discussion is developed about the estimated models and their predictive capacity.

4.2.2. Systematic Variations Model (SVM)

The model calculated (Tabla 15) that considers systematic variations in user tastes (see section 3.3) shows a substantially improved fit (see section 5). This model has kept the independent variables present in BM (section 4.2.1), with the exception of the Ticket Prices, which only appears at the interactions. Two characterisation dummy variables have also been added (Purpose: Work and Trips > 30) because in both cases the variables appear to be significant. However, the parameters of these variables are negative in both cases, which means a reduction in the perceived quality of users with these characteristics. It seems that regular customers and people that use the bus for work are more critical of the service than the other types of users.

The model has adequate consistency because it complies with all the requirements described in section 4.2.1, with one slight peculiarity. In the case where a variable, apart from appearing independently, is seen to be affected by an interaction then the condition

of non negativity needs to be fulfilled for the sum of all the parameters governing the variable. Therefore, as can be seen in the model (Tabla 15), some interactions may have a negative parameter, e.g. Age_25_35 * Comfort or Retired * Comfort. However, none of these values is strong enough to make the group of parameters negative because the parameter associated with the Comfort variable has a higher value. Although it is true that the sum of the interactions would be enough to turn the sum of the parameters negative, this cannot happen because both interactions do not occur simultaneously in reality, as a person under 25 years old cannot be retired as the average retirement age in the area where the survey took place is 65.

The model (Tabla 15) showed that Comfort continues to be the dominant influential variable, however, if the interactions are considered, the importance of Comfort could fall along with an increase in the importance of other attributes. The reasons for journey are important when evaluating certain aspects of the service, because all the possible purposes studied become significant when they interact with another variable.

Tabla 15 Artículo 1: Table 5– Systematic Variations Model (SVM)

SVM		
F.Log-Likelihood	-467.011	
AIC/N	1.331	
Count R ²	0.743	
Variable	Coef.	Test t
Constant	-4.141	-10.38
Walking time to the bus stop	0.203	3.05
Waiting Time	0.129	2.05
Travel Time	0.160	2.12
Ease of Transfer	0.211	3.05
Offered Service	0.060	0.88
Service reliability	0.177	2.41
Line Coverage	0.162	2.21
Information *	0.324	3.27
Comfort **	0.663	6.49
Driving style	0.283	4.18
Purpose: Work	-0.818	-2.18
Trips>30	-1.972	-1.96
Interactions		
Age_25_35 * Travel Time	0.510	3.05

Age_25_35 * Comfort **	-0.514	-3.06
Employed * Night / Weekend Service	0.133	2.45
Student * Ease of Transfer	0.129	2.25
Retired * Ease of Transfer	0.569	2.47
Retired * Comfort **	-0.474	-1.98
Purpose: Home * Ticket price	0.177	2.99
Purpose: Work * Line Coverage	0.427	3.08
Purpose: Study * Waiting Time	0.141	1.89
Purpose: Health * Ticket price	0.377	3.19
Purpose: Shopping * Walking time to the bus stop	0.131	1.89
Purpose: Leisure * Walking time to the bus stop	0.147	2.32
Trips_5_15* Ease of Transfer	0.071	1.95
Trips>30* Driving style	0.986	2.56
Income< 900* Walking time to the bus stop	0.080	2.21
Threshold parameters		
μ_1	1.894	16.88
μ_2	4.921	31.74

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system .

4.2.3. Systematic Variations Model with Weighted variables (SVMW)

Weighting is stated to be the inclusion of the importance that each user places on a characteristic of the public transport service. When they participated in the survey each interviewee ranked the service attributes to set an order of importance, where they then ordered from greater to lesser importance the variables defined in the survey. The weighting is derived from this ranking of importance and is included in the model as defined in section 3.3.

The consistency of the model is adequate given that it complies with the conditions of non negativity and significance mentioned in section 4.2.1, with the characteristic that the non negativity has to be fulfilled by the sum of all the parameters affecting a variable, considering both the interactions and the weighting.

The weighted variables change value depending on the importance that the users place on that particular variable. In the simply weighted variables $\delta_{nk} \cdot v_{ik}$ the variable will obtain its maximum score if the user considers it to be the most important variable within the group, or if $\delta_{nk} = 1$. However, on the contrary, if the interviewee considers the variable to be of little importance, then this variable may be omitted from the model as $\delta_{nk} = 0$. In the case of double weighting, $\delta_{nk} \cdot \delta_{mk} \cdot v_{ik}$, the importance of the variable intercedes

with the importance that the group has for the user so that for the weighting to allow the maximum scores, both the variable and the group have to be the maximum preoccupation of the user, in other words $\delta_{nk} = 1$ and $\delta_{mk} = 1$. However, if in either of the two cases the user considers that the variable is not important it is removed from the model, which occurs if $\delta_{nk} = 0$ or $\delta_{mk} = 0$.

Weightings with positive parameters cause an increase in the quality generated by that variable when the user considers it to be more important, as occurs in the case of $\delta_n * \delta_m * \text{Ticket price} * \text{Purpose: Health}$ (see Tabla 16). However, if the parameter is negative, the contribution of that variable to the Overall Quality is seen to fall at the same time as the level of importance rises.

Tabla 16 Artículo 1: Table 6 – Systematic Variations Model with Weighted variables (SVMW)

SVMW		
F.Log-Likelihood	-463.186	
AIC/N	1.323	
Count R ²	0.750	
Variable	Coef.	Test t
Constant	-4.34034	-10.84
Walking time to the bus stop	0.28815	4.2
Waiting Time	0.13341	2.11
Travel Time	0.21549	2.68
Ease of Transfer	0.2043	2.94
Offered Service	0.06723	0.99
Service reliability	0.20377	2.78
Line Coverage	0.17093	2.33
Information *	0.32107	3.25
Comfort **	0.6969	6.8
Driving style	0.31924	4.73
Purpose: Work	-0.85694	-2.3
Interactions		
Age_25_35* Travel Time	0.47365	2.83
Age_25_35* Comfort **	-0.47203	-2.82
Employed * Night / Weekends Service	0.09919	1.86
Student * Ease of Transfer	0.11383	1.98

Retired * Ease of Transfer	0.59796	2.6
Retired * Comfort **	-0.53367	-2.23
Purpose: Work * Line Coverage	0.45385	3.27
Purpose: Study * Waiting Time	0.14721	2.02
Purpose: Leisure * Walking time to the bus stop	0.15167	2.45
Trips_5_15* Ease of Transfer	0.08091	2.24
Weighted Variables		
δ_n * Walking time to the bus stop	-0.13125	-2.25
δ_n * δ_m * Travel Time	-0.16038	-2.62
δ_n * Driving style *Trips>30	0.35931	2.79
δ_n * δ_m * Ticket price * Purpose: Home	0.21639	3.28
δ_n * δ_m * Ticket price * Purpose: Health	0.45442	3.54
δ_n * δ_m * Walking time to the bus stop * Purpose: Shopping	0.42796	2.93
Threshold parameters		
μ_1	1.89808	16.83
μ_2	4.94626	31.52

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system.

4.2.4. Base Model with Random Parameters (RBM)

The inclusion of random parameters generates a model which has both constant and random parameters. In this case the parameters of the variables Ticket Price, Ease of Transfer, Information and Comfort have been defined as random. For the model to be consistent the parameters have to continue to comply with the conditions of non-negativity and significance explained in section 4.2.1, however, in this case, as these are statistical distributions both the mean and the variance of the random parameters need to be significant. The improvement obtained due to the inclusion of random parameters is discussed in greater depth in section 5.

Random parameters means show similar values to the parameters obtained in the BM model (section 4.2.1). That value can vary significantly according to the standard deviation (scale parameter for normal distributions), the Ticket Price shows the greatest standard deviation. Comfort, while being a normally distributed parameter, show a very small deviation from the mean, especially if take into account that it is the parameter with the highest value of the model.

Tabla 17 Artículo 1: Table 7– Base Model with Random Parameters (RBM)

RBM

F.Log-Likelihood	-494.197	
AIC/N	1.374	
Variable	Coef.	Test t
Constant	- 5.57038	-12.40
Walking time to the bus stop	.30743	4.32
Waiting Time	.19082	2.76
Travel Time	.24619	3.30
Offered Service	.08812	1.31
Service reliability	.22967	3.09
Night / Weekend Service	.16764	2.49
Line Coverage	.30670	4.54
Driving style	.36086	5.24
Means for random parameters		
Ticket price	.10790	1.98
Ease of Transfer	.33837	4.92
Information *	.36328	3.37
Comfort **	.75418	7.38
Scale parameters for distributions of random parameters		
Ticket price	.16334	6.67
Ease of Transfer	.19180	9.21
Information *	.18292	8.99
Comfort **	.04482	2.25
Threshold parameters		
μ_1	2.20524	13.56
μ_2	5.95623	22.63

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system.

4.2.5. Systematic and Random Variations Model (SRVM)

In the case of the Random and Systematic Variations model (Tabla 18), the differences with the not random model (section 4.2.2) resides in the fact that the interactions involving random parameters are also randomly distributed. As in the RBM, all random parameters are considered to be distributed following a Normal Distribution Function.

The consistency of this model is also adequate, complying with the conditions of negativity and significance in all cases, constant parameters, means and variances as well as in the constant and random interactions.

No-random parameters keep the same structure as in previous models. While random parameter values have been affected by the introduction of interactions. Ticket Price show a non-significant mean value, but both the deviation and heterogeneity in the mean turn out to be statistically significant, being the reason of maintaining it as a random parameter. For Ticket Price and Ease of Transfer the values of the heterogeneity are shown to be positive, while for comfort, in both interactions with young and retired people, the value of the mean is reduced substantially, lowering the importance of this attribute regarding overall quality.

Tabla 18 Artículo 1: Table 8 – Systematic and Random Variations Model (SRVM)

SRVM		
F.Log-Likelihood	-464.444	
AIC/N	1.335	
Variable	Coef.	Test t
Constant	-5.734	-11.22
Walking time to the bus stop	0.279	3.62
Waiting Time	0.219	2.89
Travel Time	0.208	2.47
Service reliability	0.239	2.92
Line Coverage	0.232	2.81
Driving style	0.393	5.12
Purpose: Work	-1.182	-2.80
Trips>30	-3.037	-1.97
Interactions		
Age_25_35* Travel Time	0.754	3.70
Employed* Night / Weekend Service	0.198	3.14
Purpose: Work * Line Coverage	0.588	3.66
Purpose: Study * Waiting Time	0.171	1.73
Purpose: Shopping*Walking time to the bus stop	0.183	2.26
Purpose: Leisure * Walking time to the bus stop	0.194	2.50
Trips>30* Driving style	1.519	2.70
Income< 900* Walking time to the bus stop	0.112	2.56

Means for random parameters		
Ticket price	0.010	0.15
Ease of Transfer	0.287	3.47
Information *	0.454	3.85
Comfort **	0.969	8.02
Scale parameters for distributions of random parameters		
Ticket price	0.187	7.09
Ease of Transfer	0.259	10.96
Information *	0.166	7.68
Comfort **	0.112	5.09
Heterogeneity in the means of random parameters		
Purpose: Home*Ticket price	0.218	2.94
Purpose: Health*Ticket price	0.516	3.42
Student * Ease of Transfer	0.199	2.81
Retired * Ease of Transfer	0.845	3.12
Trips_5_15* Ease of Transfer	0.103	2.34
Retired * Comfort **	-0.729	-2.57
Age_25_35* Comfort **	-0.768	-3.66
Threshold parameters		
μ_1	2.592	12.88
μ_2	6.895	20.42

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system.

4.2.6. Systematic and Random Variations Model with Weighted variables (SRVMW)

This final model considers both the systematic and random variations in user taste as well as the weighting of the variables according to the importance each user associates with them. This represents the most complex model estimated in this study. The complete model can be seen in Tabla 19.

The consistency of the model is adequate and complies with the statements in section 4.2.1 considering the group of parameters affecting each one of the variables.

The weights of the variables vary greatly as they consider all the effects present in the model which means that the analysis of the model is also a complex task. Generally, most of the parameters are seen to be not considered as random, in particular those that suffer some kind of weighting. The improvements resulting from the use of this model are discussed in greater detail in section 5.

Weighting of the variables have changed the structure of the random parameters. The parameter of the weighted variable δn^* Walking time to the bus stop is now considered as random. Ticket Price becomes no-significant in an independent way, being only significant weighted and interacting with Home and Health Purposes. Rest of parameters maintain the structure of previous models.

Tabla 19 Artículo 1: Table 9 – Systematic and Random Variations Model with Weighted variables (SRVMW)

SRVMW		
F.Log-Likelihood	-461.817	
AIC/N	1.327	
Variable	Coef.	Test t
Constant	-5.73758	-11.69
Walking time to the bus stop	0.35827	4.66
Waiting Time	0.1807	2.4
Travel Time	0.26199	2.99
Offered Service	0.09707	1.35
Service reliability	0.27713	3.4
Line Coverage	0.21433	2.6
Driving style	0.40202	5.47
Purpose: Work	-1.06449	-2.58
Interactions		
Age_25_35* Travel Time	0.62038	3.11
Employed* Night / Weekends Service	0.13404	2.29
Purpose: Work * Line Coverage	0.55626	3.55
Purpose: Study* Waiting Time	0.16976	1.77
Purpose: Leisure * Walking time to the bus stop	0.19402	2.64
Weighted Variables		
$\delta n^* \delta m^*$ Travel Time	-0.21406	-2.9
δn^* Driving style *Trips>30	0.45992	2.78
$\delta n^* \delta m^*$ Ticket price * Purpose: Home	0.27849	3.78
$\delta n^* \delta m^*$ Ticket price * Purpose: Health	0.61475	4.02
$\delta n^* \delta m^*$ Walking time to the bus stop * Purpose: Shopping	0.56547	3.11

Means for random parameters		
Ease of Transfer	0.25988	3.21
Information *	0.45028	3.89
Comfort **	0.94291	7.84
δn^* Walking time to the bus stop	-0.15761	-2.27
Scale parameters for distributions of random parameters		
Ease of Transfer	0.21199	9.61
Information *	0.14178	6.9
Comfort **	0.16919	7.67
δn^* Walking time to the bus stop	0.11354	3.42
Heterogeneity in the means of random parameters		
Student* Ease of Transfer	0.16848	2.39
Retired * Ease of Transfer	0.79904	2.95
Trips_5_15* Ease of Transfer	0.1106	2.63
Retired* Comfort **	-0.73297	-2.56
Age_25_35* Comfort **	-0.61993	-3.02
Threshold parameters		
μ_1	2.45654	12.85
μ_2	6.50259	21.4

*Average of the three variables within the group.

**Average between Comfort of the buses and Air conditioning / heating system.

5. DISCUSSION

A comparison will now be discussed about all the models estimated in this study (Tabla 20). The comparison was done by calculating the following indicators: AIC/N (Akaike, 1973), log-likelihood and *Count R²* (William H Greene and Hensher, 2010; Scott and Freese, 2006). The values of AIC/N (Akaike, 1973) and Log-likelihood are useful for comparing different models and to see which is best, however, they do not show the real representational capacity of the model. In the case of the AIC/N value, a model fits better than another the lower the value of the indicator is, in the case of Log-likelihood this occurs when the value is closest to 0. The *Count R²* is much more representative and instinctive about the real predictive capabilities of a model. This can be calculated by comparing the predictions made by the model with real observations. For example, a *Count R²* = 0.724 shows that the model gets the right result 72.4% of the time. In the case of models with random parameters, the calculation of this indicator is not a simple

task so the comparison is made by comparing the other two indicators without considering random parameters.

Tabla 20 Artículo 1: Table 10 –Comparison of Models

Model	Count R²	Log-Likelihood	AIC/N
BM	0.724	-497.064	1.371
SVM	0.743	-467.011	1.331
SVMW	0.750	-464.097	1.323
RBM	N/A	-494.197	1.374
SRVM	N/A	-464.444	1.335
SRVMW	N/A	-461.817	1.327

It needs to be pointed out first that the fits of the models calculated in this study are adequate, even though there is an improvement when more complicated models are estimated the simplest model (BM) still has good predictive capabilities. An analysis of the changing fits when various modifications are introduced shows that the best improvement occurs when systematic variations are included in the model. This occurs for the ordinary models and the random models and agrees with other studies showing the need to analyse user heterogeneity when evaluating quality (Bordagaray et al., 2014; dell’Olio et al., 2010). Similarly, when considering random variations in taste as well as systematic heterogeneity (SRVM model), the log-likelihood value also shows us there is an improvement. Nevertheless, the AIC/N value worsens because this indicator is dependent on a number of variables, resulting in a worse value when more variables are involved, as occurs when random parameters are included in models.

The weighting of the variables is seen to generate an improved fit, however, the improvement is not as substantial as that resulting from the inclusion of user heterogeneity.

Finally, it is worth mentioning that the increased complexity of the models has resulted in better fits and the model with the best fit of them all is the one which included a combination of both systematic and random variations in user taste with the weighting of the variables (SRVMW). This model was also the most complex of the models estimated in this study.

6. CONCLUSIONS

This article has described a methodology developed for analysing the perceived quality of users of a public transport system. Knowledge of the attributes which have a greater or lesser influence on overall service quality represents a useful tool for administrations and operators of these services as it allows them to concentrate their investments where they are really needed.

Ordered probit models have also been shown to be able to convincingly model user perceived quality, especially the non linearity present in the evaluations received in the surveys, given that one of the benefits of modelling with ordered models resides in the modelling of the non linearity of the evaluation categories. What this means is that it is not the same thing to go from a score of “Bad” to one of “Normal, than it is to go from one of “Good” to “Very Good”. This effect can be tested in all the models estimated in this study by looking at the limit parameter μ .

It needs to be highlighted that the improved fit of the models resulting from the weighting of the variables was not particularly substantial; the most noteworthy improvement came from the inclusion of user heterogeneity. Considering that when the surveys were being asked the most laborious and costly part was the task involved ordering the variables from greater to lesser importance, we can say that for a practical application where a company may wish to know the quality the users perceive of their services then weighting the variables would not be recommended because it was not shown to be a particularly efficient activity. On the contrary, as the characterisation of users is standard practice in this type of survey, it would be highly recommendable to include user heterogeneity in the models. However, although ranking the variables was not shown to be an overly efficient practise for modelling it did provide useful information for policy design as it revealed the comparative importance of certain variables over others.

Although this study has provided some satisfactory results, it leaves many other questions in the air. Firstly, it would be useful to establish other criteria for weighting the variables, because the methodology followed here was possibly not adequate in that it may not have provided the best results. Furthermore, this case has proposed a combined model for all the analysed bus lines which means it would be interesting to analyse the lines separately considering their heterogeneity, a concept that has already been demonstrated (dell'Olio et al., 2010). Finally, a step forward in this type of study would be to create a system of continuous control which would improve the study of quality by enabling the analysis of how it changes over time (de Oña et al., 2016c).

ACKNOWLEDGEMENTS

The Authors acknowledge the support of the Ministry of the Economy and Competition of the Spanish Government through the project TRA2015-69903-R.

4.2 ARTÍCULO 2: MODELLING USER SATISFACTION IN PUBLIC TRANSPORT SYSTEMS CONSIDERING MISSING INFORMATION

4.2.1 Indicadores bibliométricos de la revista de publicación del artículo 2 en el JCR



Título de la revista: Transportation

ISSN: 0049-4488

eISSN: 1572-9435

Editorial: Springer Nature

País de publicación: Países bajos

Volumen de publicación: Publicado online

Fecha de publicación: 28 de marzo de 2019

Páginas: N/A

DOI: <https://doi.org/10.1007/s11116-019-09996-4>

Factor de impacto JCR

Tabla 21 Factores de impacto de la revista del artículo 2

Año de publicación (2018) ²	3,457
Promedio de 5 años	3,851

Ranking

Tabla 22 Ranking de la revista del artículo 2

Área de conocimiento	Posición	Percentil	Cuartil
Transportation	8/36	79,467	Q1

Numero de citas que posee el artículo

Tabla 23 Número de citas del artículo 2

Plataforma de medición	Número de citas
Google Scholar	4
Web of Science	0
Scopus	0

² Se considerará como el valor del factor de impacto el establecido para el año de publicación del artículo. En el caso que el valor del año correspondiente no esté disponible se considerará el último año disponible.

4.2.2 Transcripción del artículo 2: Modelling user satisfaction in public transport systems considering missing information

MODELLING USER SATISFACTION IN PUBLIC TRANSPORT SYSTEMS CONSIDERING MISSING INFORMATION

Eneko Echaniz

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

Chinh Ho

The University of Sydney, Institute of Transport and Logistics Studies, The University of Sydney Business School, 378 Abercrombie St, Darlington NSW 2008 (Australia)

Andres Rodriguez

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

Luigi dell'Olio

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

ABSTRACT

Collecting data to obtain insights into customer satisfaction with public transport services is very time-consuming and costly. Many factors such as service frequency, reliability and comfort during the trip have been found important drivers of customer satisfaction. Consequently, customer satisfaction surveys are quite lengthy, resulting in many interviews not being completed within the aboard time of the passengers/respondents. This paper questions as to whether it is possible to reduce the amount of information collected without a compromise on insights. To address this research question, we conduct a comparative analysis of different Ordered Probit models: one with a full list of attributes vs. one with partial set of attributes. For the latter, missing information was imputed using three different methods that are based on *modes*, single imputations using predictive models and multiple imputation. Estimation results show that the partial model using the Multiple Imputation method behaves in a similar way to the model that is based on the full survey. This finding opens an opportunity to reduce interview time which is critical for most customer satisfaction surveys.

Keywords: Missing information, multiple imputation, user satisfaction, ordered probit, perceived quality.

1. INTRODUCTION

Research on the perceived quality or the satisfaction of the users usually relies on customer satisfaction surveys conducted using a revealed preference survey method. Data collection is usually the most time-consuming and costly part, especially when a face-to-face survey method is used. While this survey method undoubtedly delivers the high data quality, its completion/response rate depends heavily on the interview duration with lengthy questionnaire resulting in a lower response/completion rate. Thus, finding a way to shorten the survey length would improve the effectiveness of customer satisfaction studies. This article proposes a way to do so through a comparative analysis of different models. These models are based on data from customer satisfaction surveys with full and partial list of attributes. Partial dataset is obtained after randomly deleting half of the information available in the original survey. No statistical difference between the two methods will mean that it is possible to reduce the amount of data collected in customer satisfaction surveys. To this end, missing data are imputed using three different methods in order to identify the most adequate method for imputing non-collected information. The first method uses the “mode” of each attribute to fill out the data for respondents who were not shown these attributes. The second uses Ordered Probit models for each attribute and the final method uses Multiple Imputation Process. Different Ordered Probit models are then estimated for the different databases and results compared to check if the models obtained with the partial information databases are correlated with the model based on the complete database.

The remaining of this paper includes 7 sections. The next section summarises a general view of the state of the art regarding the study of satisfaction in public transport systems, reviewing the most relevant studies. Methodology is described in section 3 with analysis results presented in section 4. Section 5 discusses the important findings and identify areas for future research.

2. LITERATURE REVIEW

Satisfaction surveys have proved to be a reliable and robust method to measure the users' perceived quality of public transport systems. Many studies contributed to this fast-growing literature, from a generic analysis of perceived quality (Parasuraman et al., 1985) to the more advanced method that focuses on the provision of public transport services (dell'Olio et al., 2010; dell'Olio et al., 2011; Felleson and Friman, 2008; Rojo et al., 2013; Wongwiriya et al., 2017).

Most of these studies have focused on identifying key drivers/attributes of the transport system that best describe public transport services. Examples are the Quattro project (EC, 1999) which used eight sets of attributes, or the work of Hensher et al. (2003) that employed Service Quality Index (SQI). Another line of research in this literature focused on improving the method used for modelling the data collected. A variety of modelling methods have been used such as basic statistics (Eboli and Mazzulla, 2011), Ordered Data Models (Bordagaray et al. 2014; dell'Olio et al. 2010; Echaniz et al. 2017), structural equations model (Das et al., 2017; de Oña et al., 2013; Rahman et al., 2016) and decision tree (de Oña et al., 2016d; Hernandez et al., 2016; Machado-León et al., 2017; Tsami and Nathanail, 2017). Recently, we see some exceptions that aim to optimize the data collection. Typical examples are Rose and Bliemer (Rose and Bliemer, 2014, 2013, 2009)

where efficient stated preference S-design is used to minimize the sample size; however, similar efforts in optimising surveys are not observed in the revealed preference domain.

The data collection process is essential part of any customer satisfaction study that usually use on-board intercept followed by face to face interviews (Bordagaray et al. 2014; dell'Olio et al. 2010; Echaniz et al. 2017) or self-administered questionnaire accessible via QR codes or URL links provided at intercept at public transport stops/stations where the passengers board or alight (Guirao et al., 2015). As for customer satisfaction surveys, the survey duration is a key factor for obtaining valid and quality responses. A long questionnaire generates rich data for the subsequent analysis but this significantly reduces the response/completion rate, resulting in fewer samples for a given budget. Conversely, short surveys can improve sample size at the expense of less data being collected such that statistical model results are not very reliable and robust, since the model cannot control for some important factors that were not collected in the survey. Trade-offs between data richness and budget depends directly on the target sample size, the duration of the survey and the survey method (face to face, online, or app). The aim of this study is to obtain robust models, not by reducing the number of observations, but reducing the amount of data required from each respondent, in other words, reducing the time required to complete each survey. The benefit of reducing surveying time increases as the sample size becomes larger.

The literature shows that user satisfaction studies usually require a long list of factors that requires respondent's feedback, with the data collection lasting from several weeks to even a few months. In Rahman et al., (2016) for example, surveys were conducted to 2008 public transport users during the months of June and July 2015, with a survey consisting of two sections, one for obtaining socio-economic data and another for obtaining the satisfaction of 21 attributes of the system. In Rissel et al., (2016) a total of 512 online surveys were conducted during the months of September and October to obtain information on the mode of transport used by users and the level of satisfaction they had with it. In Guirao et al., (2016) 850 face-to-face surveys were carried out, of which 813 were valid complete answers. The length of the respondent period was 2 weeks. In Tyrinopoulos and Antoniou (2008) 1474 survey where set as minimum to have a sample representative enough for their study, where a total of 5 companies where evaluated. In St-Louis et al. (2014) an online approach was taken. The invitation to participate in the survey was sent via email targeting university staff and students. The response rate was 31.7%, 3377 complete responses from the 20,851 invitation sent. Participation was incentivised by different prizes and all respondents received a reminder 2 weeks after the first email was sent. At the end, the survey was kept active for 35 days during March and April 2013. In Abenoza et al., (2017) unlike the previous ones, there was a very extensive database obtained by the Swedish Public Transport Barometer for the years between 2001 and 2014 with about 450,000 useful telephone surveys. The aforementioned studies are only a small example of satisfaction studies carried out in the last few years. Hence, it can be seen that public transport satisfaction studies require the completion of a large number of surveys. Therefore, an improvement in the efficiency of this process would considerably improve the total cost of the entire process, as long as the quality of the data allows subsequent analyses.

Regarding the modelling methodology, several studies have shown (Bordagaray et al. 2014; dell'Olio et al. 2010; Dell'Olio et al. 2011; Echaniz et al. 2017; Rojo et al. 2013) that ordered data models are very adequate to model customer satisfaction with public transport services. These models requires a series of very specific data, which are composed by a dependent variable, the overall satisfaction of the service, and independent variables, the attributes of the service. Each respondent must evaluate all the variables, which means that for a survey in which 24 attributes are used to define the system, the respondent must answer at least 25 questions (24 attributes and the overall satisfaction), in addition to background questions relating to individual characteristics such as age and gender. These need of complete data observations have been the main reasons for choosing this modelling method for this study.

To analyse a missing database as if all the data were available, it is necessary to establish a methodology to fill in the missing information. The statistical processes relating to the missing information have evolved considerably in recent years. It has been a subject of many sociology and psychology studies, of which the work of Schafer and Graham (2002) stands out. The authors provide an extensive review of the state of the art regarding the types of missing data and the different imputation methods available. There are several types of missing data according to the nature of the reason why they are missing, according to the classification rooted to Rubin (1976).

In order to classify the current study within this classification it is necessary to understand that the missing information has not been due to a decision of the respondent not to answer a question, but due to the design. That is, part of the available information has been deliberately eliminated to form a reduced survey with fewer questions. Since the elimination of the data has followed a random criterion, the nature of the missingness does not depend on any of the variables belonging to the survey, being a case defined as "missing completely at random" MCAR. In (Graham et al., 1996) this type of scenarios is defined as Planned Missing Value patterns, in other words, survey was intentionally planned to have missing information. When the type of missing information responds to an MCAR nature, simple methods can be applied (Donders et al., 2006). However, in this study we find ourselves in a special MCAR situation, where all the observations have missing data, and therefore there are techniques like listwise deletion (deletion of observations with missing data) that cannot be used. However, within these simple methods (Donders et al., 2006), or also called older methods (Schafer and Graham, 2002), there are several suitable methods for the case that concerns us. On the one hand, there is the possibility of replacing the missing data with the means of the available observations of each variable, in our particular case, as we are dealing with discrete qualitative responses it has been considered more accurate to replace the missing data with the most common response (mode) of that variable. On the other hand, it is possible to apply the single imputation method, based on imputing a single value for each missing data, filling it with a plausible value. This imputation is made by inferring a value for the missing data based on the information that we have available. Although this simple methods can give acceptable results, several studies recommend using more sophisticated methods based on Maximum likelihood (Graham et al., 1996) or Multiple imputation (Donders et al., 2006; Graham et al., 1996). More explicitly, in Donders et al. (Donders et al., 2006) it is demonstrated that the use of the multiple imputation (MI) approach leads to results with

correct standard errors, especially in situations where missing data is MCAR. In the same way, Graham and Schafer (1999) showed that MI performs very well in small samples even with as much as 50% missing data.

The MI method was initially developed by Donald Rubin (Rubin, 1977) and has proven to be a very effective method to obtain missing data in non-responses. Indeed, MI method is very popular in social sciences and medical studies. Examples in the field of medicine include (Burton et al., 2007; Newgard et al., 2018; Pettersson et al., 2018; Sterne et al., 2009; Troyanskaya et al., 2001; van Buuren et al., 1999) and (Alegria et al., 2004; Allison, 2000; König et al., 2018; Love et al., 2018; Phan et al., 2016; Roth, 1994; Zou, 2015) for social sciences. Very few applications have been found in transport research with exceptions being (Chiou et al., 2014; Henrickson et al., 2015; Li et al., 2015; Tang et al., 2015) that uses MI method to fill the missing data in traffic flows or loop detectors.

Thus, to the best of authors' knowledge, previous studies have mainly focused on filling missing information caused by problems during the surveying process i.e.: non-responses. In this specific case, we propose the possibility of using the same methodology to verify that it is possible to obtain similar results considering a partial sample, similar to what was proposed in (Graham et al., 1996). Moreover, this technique has not been yet applied for the specific case of users' satisfaction in transit services.

3. MODEL SPECIFICATIONS

3.1. Ordered Probit Modelling

The Ordered Probit model was first proposed by McKelvey and Zavoina, 1975, 1971 for the analysis of choices and ordered, categorized or non-quantitative responses.

The ordered data models are based on dividing a continuous utility space (users' satisfaction in this case) in discrete bands through a system of limitations (Greene y Hensher, 2010).

$$\begin{aligned}
 q_i^* &= \beta' x_i + \varepsilon_i, \quad i = 1, \dots, n, \\
 q_i &= 1 \text{ si } \mu_0 < q_i^* \leq \mu_1 \\
 &= 2 \text{ si } \mu_1 < q_i^* \leq \mu_2 \\
 &= 3 \text{ si } \mu_2 < q_i^* \leq \mu_3 \\
 &= \dots \\
 &= J \text{ si } \mu_{J-1} < q_i^* \leq \mu_J
 \end{aligned} \tag{45}$$

The key idea of the model is that the observations made are not a simple accumulation of discrete results that can be ordered in a certain way, but consist of a transformation of a single continuous variable that must be ordered.

The model contains the unknown marginal utilities, β , in addition to $J + 2$ threshold parameters μ_j , all of them to be estimated by n observations. The data consists on the variables x_i of each observation and the resulting observation q_i of each one of them. The random variable ε_i completes the model. It is assumed that the random variable ε_i is distributed according to a known distribution function and defined throughout the real domain. Focusing the models on the problem raised in this study. Let's suppose a series of answers available for each of the respondents, where the options are the following:

0 Very Bad

1 Bad

2 Normal

3 Good

4 Very Good

The regression model shows an underlying and at the same time not observable preference on the evaluated question, q_i^* . Each individual surveyed does not provide the value of q_i^* , but a limited version of it divided into five possible options, one of which is closest to his exact preference. The probabilities associated with the observed responses are:

$$\text{Prob}[q_i = j | x_i] = \text{Prob}[\varepsilon_i \leq \mu_j - \beta'x_i] - \text{Prob}[\mu_{j-1} - \beta'x_i], \quad j = 0, 1, \dots, J \quad (46)$$

The established model describes the probability of occurrence of the values of the results. It does not describe a direct relationship between the evaluation q_i and the parameters x_i , because there is no obvious regression relationship between both parameters, since q_i is mere a label.

For the estimation of the parameters it is necessary to establish a series of normalizations. First, to keep the positive signs for all probabilities, it is necessary that $\mu_j > \mu_{j-1}$. Second, if the model must exist in the complete real domain, then $\mu_{-1} = -\infty$ y $\mu_J = +\infty$. Since the data does not contain unconditional information about the scale of the dependent variable (in case of modifying the scale of q_i^* with any positive value, modifying the scale of the unknown values μ_j y β with the same value the characteristics of the observations will remain the same) it is not possible to estimate the free variance parameter $\text{Var}[\varepsilon_i] = \sigma_\varepsilon^2$. It is advisable to make a restriction based on $\sigma_\varepsilon = \text{const.}$ It is usual to assume variance equal to one in the case of a Probit model and variance equal to $\pi^2/3$ in the case of Logit. Finally, assuming that $\beta'x_i$ has a constant term, it is necessary to set $\mu_0 = 0$. The calculation of the parameters of the models is done by a maximum likelihood estimation (Greene, 2007, 2008; Pratt, 1981), which equation to maximize is:

$$\log L = \sum_{i=0}^n \sum_{j=0}^J m_{ij} \log[F(\mu_j - \beta'x_i) - F(\mu_{j-1} - \beta'x_i)] \quad (47)$$

3.2. Multiple Imputation for missing information

The goal of the multiple imputation is to complete the missing data, in a way that the resulting data can be statistically analysed and modelled in a similar way to the complete database. The theoretical foundation on which the multiple imputation is based on is the repetitive imputation (Rubin, 2004, 1996, 1977). This means that for each missing data value m values (as opposed to 1) are imputed. Considering the fact that the missing data have been eliminated randomly, it can be said that the missing data corresponds to a MCAR type, so the use of this method is appropriate (Donders et al., 2006).

The methodology used to perform the multiple imputation is called the Fully Conditional Specification (FCS), which uses an iterative Monte Carlo method with Markov chains (van Buuren, 2007). The FCS approach is based on variable-by-variable imputation of data, specifying an estimation model for each one of the variables with missing data. The FCS tries to define $P(X, C, R | \theta)$ by specifying a conditional density $P(X_i | C, X_{-i}, R, \theta_i)$

for each X_i , this density is used to impute X_i^{mis} given some C , X_{-i} and R . An imputation consists of a complete cycle through all X_i (van Buuren, 2007). Where X represents the evaluation of the attributes, C the characterization variables, θ the parameters of the imputation model and R an indicator that show if X is a missing or observed value. The imputation is made by using the Gibbs sampling methodology (Casella et al., 2016; Gilks et al., 1996) assuming that the conditional density distribution exists. This methodology has been used in a large number of simulation studies (Brand, 1999; Brand et al., 2003; Horton et al., 2016; Raghunathan et al., 2001; Van Buuren et al., 2006) that have provided sufficient evidence that the results obtained through the FCS are generally unbiased and have adequate coverage.

In order to optimize the imputation process it has been assumed that the satisfaction data is a scale type variable with values between 0 and 4, so the imputation model follows a linear regression methodology rounding to the nearest whole value. This has enabled the imputed values to match the actual values of the data. It has been proven that the predictive mean matching procedure, a variant of the linear regression that equals the imputed values with the closest observed value, generates worse results. In Graham et al. (2007) they recommend a high number of imputations for these types of cases. However it has been empirically verified that, for this specific practical case, with 5 imputations the results are acceptable enough. So it has been decided to maintain this number of imputations (5) mainly for efficiency reasons. For the regression model, Y_j corresponds with the attributes with missing evaluations and X with all the socioeconomic variables (Tabla 24) plus the overall satisfaction of the service.

3.3. Comparison

According to the final objective of this study, in which it is intended to analyse if it is possible to obtain similar results based on a partial information database, 3 methodologies are proposed to perform the modelling.

The starting point is the model that it will be called BASE, which is estimated considering the complete database. For the rest of the models, half of the satisfaction data have been randomly eliminated, that is, creating a hypothetical scenario where only 12 of the 24 attributes would have been answered by the respondents. For modelling, missing information need to be fulfilled so 3 different methods have been used to achieve that.

The first method is based on using the “mode” of the answers to complete the missing information of each attribute. That is, to use the most common value among the respondents for each attribute. In other words, the satisfaction of a user who does not make the evaluation of an attribute will be equal to the most commonly chosen value by those that did evaluate it. This model will be called MODE throughout the rest of the article.

The second method consists of estimating J Ordered Probit models, one for each of the attributes. This way, a missing satisfaction value is imputed from a model estimated with the existing responses for that attribute, based on the socioeconomic characteristics of the people who have evaluated it. Being explained as:

$$y_{ij}^* = \delta_{ij} \beta' x_i + \varepsilon_i, \quad i = 1, \dots, n \text{ and } j = 1, \dots, 24 \quad (48)$$

Where, y_{ji}^* represents each one of the 24 evaluated attributes and x_i the different socioeconomic variables; δ_{ji} obtains the value 1 if the attribute j is evaluated by the respondent i and 0 otherwise, up to a maximum of $\sum_{j=1}^{24} \delta_{ij} = 12$ per respondent, since it has been assumed that in the restricted version of the survey the respondents would only perform half of the evaluation exercises. This model will be called ATTRIBUTE throughout the article.

Finally, the last method used to complete the missing data have been through the use of the multiple imputation procedure (section 3.2). As indicators to infer the missing data, both the socioeconomic variables and the evaluations made to all the attributes have been used, as well as the overall satisfaction of the service. A total of 5 imputations have been carried out with 100 interactions each. The results of the Multiple Imputation consist in the generation of 5 new databases, 1 for each imputation. In order to obtain a single model, an OP model have been estimated for each one of these databases and then the average of the parameters have been use for the comparison. This model will be referred by the acronym MI.

4. CASE STUDY

4.1. Satisfaction survey

The data used in this study was obtained from a satisfaction survey carried out in 2015 in the city of Santander, a small-medium size coastal city located in the north of Spain. At the time when the survey was conducted, the city had around 173,000 inhabitants with the metropolitan area reaching 240,000 residents. Buses are the only public transport of the city. The bus network has 22 lines, of which 16 were surveyed in this study.

Field surveys were conducted over 15 working days in the months of April and May 2015. The surveys were carried out on board using a face to face method. In case the survey could not be completed during the respondent's journey, the interviewer had two options: either leaving the bus with the respondent and finishing the survey at the stop or discarding the survey and find another respondent on board. If the respondent chose the former option, they then wait for the next bus to come, and then continue on-board recruitments and interviews. In both cases the efficiency of the survey process was affected. The minimum sample size n was set in 700 completed surveys, being calculated by using equation (5) (Bordagaray et al. 2014; dell'Olio et al. 2010; Echaniz et al. 2018). For which, the most conservative value was taken: $p = 0.5$. In the end, a total of 747 complete observations were obtained with a ratio of approximately 4 complete surveys per hour per interviewer.

$$n \geq \frac{p(1-p)}{\left(\frac{e}{z}\right)^2 + \frac{p(1-p)}{N}} \quad (49)$$

Where:

e : standard error of the parameter estimate

z : critical z – statistic value ofr the desired level of confience (eg 95%)

N : Number of passengers at rush hour

p: probability associated with the choice

The survey included two main parts with the collected information summarised in Tabla 24. The first part seeks the respondent's socioeconomic characteristics and usage of public transport services. The second part focuses on user overall satisfaction with public transport service (OS) and on a subset of attributes that represent different aspects of the service. Level of satisfaction was measured using a 5 point Likert scale.

Tabla 24 Artículo 2: Table 1 - Variables included in the survey

Characterization	PT Service attributes
Gender	Access time to the stop (AT)
Age	Waiting time at the stop (WT)
Work status	On board travel time (TT)
Driving license ownership	Egress time, last stop to final destination (TD)
Car ownership	price/fare (PR)
Trip purpose	Ease of transfer (TR)
Frequency of use	Service frequencies (SE)
Usual payment system	Service reliability/punctuality (SR)
Monthly income	Special lines (EL)
	Night services / weekend services (NS)
	Line coverage (LC)
	Information at stops (IS)
	Information on webpage and mobile platforms (IWM)
	Information on board (IB)
	Occupancy level (OC)
	Air conditioned vehicles (CA)
	Priority seats for people with disability (RM)
	Comfort of the buses (CM)
	Cleanliness (CL)
	Possibility to carry large objects (OB)
	Driving style (DS)
	Driver kindness (DK)
	Hybrid buses (HY)

Noise (NO)

The sample was made up of 71% women, who are over-represented. Two thirds of the respondents are under 44 years old and nearly half working (49%) with a further quarter studying part-time or full-time. About six in ten respondents having a driving license (59%) but only four in ten own a car. Regarding the use of public transport service, regular users, defined as those using bus services between 5 and 15 times per week, accounts for nearly half of the sample where the main reason for bus travelling is commuting (work or study). The vast majority of the respondents use contactless card with cash payment accounting for only 5% of the sample. Regarding personal income level, a majority of the respondents have low to medium income levels, with high income respondents accounting for only 8%. A little more than the third part of the respondents (38%) preferred not to answer this question, a usual result since it is a very sensitive question. Tabla 25 provides a summary of the respondents.

Tabla 25 Artículo 2: Table 2 - Descriptive analysis

Gender	Male	29%
	Female	71%
Age	<25	28%
	25-34	16%
	35-44	22%
	45-54	16%
	55-64	11%
	>65	7%
Work status	Worker	49%
	Unemployed	17%
	Student	25%
	Retired	9%
Driving license	Yes	59%
	No	41%
Car ownership	Yes	40%
	No	60%
Frequency of use	< 5 trips/week	29%
	5 - 15 trips/week	50%
	15 - 30 trips/week	18%

	> 30 trips/week	3%
Trip purpose	Home	32%
	Work	22%
	Study	13%
	Health	4%
	Shopping	7%
	Leisure	11%
	Other	11%
Payment system	Contactless transport card	95%
	Cash	5%
Monthly income	Low (< 900€)	31%
	Medium (900 - 1500€)	23%
	High (1500 - 2500€)	7%
	Very high (> 2500€)	1%
	Unknown (not reported)	38%

Regarding user satisfaction with the public transport service, Figura 4 shows the level of satisfaction of respondents. For brevity, the 5-point Likert scales are coded from 0 to 4, with 0 being "Very Bad" and 4 being "Very good". Additionally, since the aim of the study is to analyse the possibility of obtaining similar results based on a reduced data base. A comparison was made comparing, on one hand, the average satisfactions obtained for the different attributes through the complete database. And on the other hand, the average value of the satisfaction obtained after eliminating half of the available information (partial database), just as it has been done for the modelling process (section 3.3).

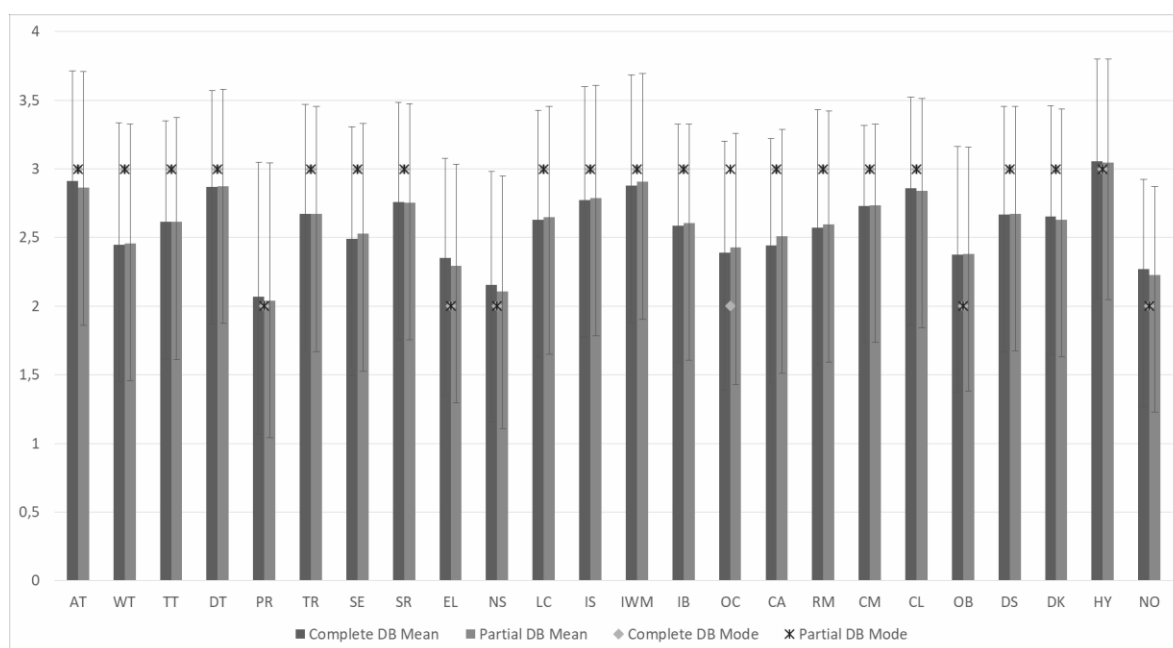


Figura 4 Artículo 2: Figure 1 – Users' satisfaction

The results show that users are generally satisfied with the service and with all aspects that describe the services they use. The attribute that is considered as worst is the fare. This fact can be understood as a strategic response, since the users do not tend to evaluate this attribute well for fear of a possible increase of the service fares, still, the average value shows that it is not considered as an unsatisfactory factor for users. On the contrary, an attribute that valued the most is the use of hybrid buses. Any action associated with an environmental improvement of the service is generally considered good by the users.

The comparison made between the two databases shows that, even having half of the information, the average difference in satisfaction level between the two datasets is small, with differences in means being less than 3% in all cases. The biggest difference in the mode is found in the occupation attribute, where the mode changes from a "normal" evaluation (value 2) to a "Good" evaluation (value 3). One possible reason for this would be a random elimination of the attributes, the worse attributes may have been eliminated. However, this difference is only shown in one variable of the whole set of attributes so it can be considered an outlier. The standard deviations also show small differences, usually less than 3% with the exception of the Access time to the stops (AT) which shows a variation close to 6%. We can safely conclude that based on the results shown in Figura 4 the average level of satisfaction are very similar between the two datasets.

4.2. Modelling results

Four Ordered Probit models were developed. One model was estimated using the complete dataset and this model is referred to as BASE. The remaining three models are developed from the partial dataset obtained after randomly deleting 50% of the evaluations made in the original survey, which resembles a hypothetical survey where only half of the attributes would have been answered by each respondent. Missing information was imputed using the three methods described in section 3.4.

The attributes included in each model have been selected following a step-by-step process until the resulting parameters have the correct sign (positive sign except for the constant that must be negative (Echaniz et al., 2018)) and are statistically significant. This can be seen in Tabla 26 where t-values are included in parentheses. Significant parameters are shown in bold (at least at a 10% level) so that similarities and differences between models can be spotted easier. The significant parameters are largely similar between different models. The most similar model to BASE in terms of the significance level of the parameters is the MODE model, where 79% of the 24 parameters have the same level of significance compared to BASE. The model calculated using the database fulfilled by attribute specific models (ATTRIBUTE) shows a lower correlation, with 71% of the parameters showing a similar significance level. Finally, the MI model, derived from Multiple Imputation, lie in the middle with a coincidence of 75% of the parameters.

Only two threshold parameters are shown in the models because there was no “Very Bad” evaluation observed in the survey for the dependent variable (OS). Instead, the value 0 now represents the grouping of the "Very Bad" and "Bad" responses.

Tabla 26 Artículo 2: Table 3 - Ordered probit models: parameter estimates and t-values

Parameters	Base	Mode	Attribute	MI ³
Constant	-4.37 (-11.22)	-7.81 (-11.89)	-8.47 (-15.25)	-3.81 (-10.82)
Access time to the stop	0.23 (3.71)	0.26 (3.35)	0.35 (4.37)	0.27 (5.05)
Waiting time at the stop	0.13 (2.14)	0.18 (2.59)	0.36 (4.73)	0.22 (4.37)
On board travel time	0.18 (2.47)	0.25 (3.11)	0.39 (4.52)	0.16 (3.37)
Egress time	0.12 (1.65)	0.24 (2.56)	-	-
Price/fare	0.07 (1.37)	0.09 (1.34)	0.07 (1.14)	-
Ease of transfer	0.26 (4.2)	0.31 (3.93)	0.36 (4.23)	0.24 (4.07)
Service frequencies	0.09 (1.29)	0.29 (3.88)	0.13 (1.53)	0.19 (3.21)
Service reliability	0.19 (2.69)	0.30 (3.49)	0.30 (3.26)	0.02 (1.7)
Special lines	-	-	-	0.1 (2.86)
Night/weekend services	0.14 (2.38)	-	0.22 (2.78)	0.12 (2.35)
Line coverage	0.25 (3.92)	0.32 (4.2)	0.31 (3.91)	0.14 (2.94)
Information at stops	0.11 (1.78)	0.22 (2.82)	0.19 (2.27)	0.02 (2.44)
Information on webpage and Mobile platforms	-	0.19 (2.32)	0.27 (3.28)	0.08 (1.64)

³ Average of the values obtained in the 5 models based on Multiple Imputation

Information on board	-	-	0.31 (3.56)	-
Occupancy level	-	-	0.18 (2.41)	0.03 (2.85)
Air conditioned vehicles	0.26 (3.99)	0.16 (2.14)	0.37 (4.52)	0.19 (3.58)
Priority seats for people with disability	-	-	0.24 (3.09)	0.04 (4.2)
Comfort of the buses	0.37 (4.29)	0.30 (2.99)	-	0.42 (5.57)
Cleanliness	-	-	-	-
Possibility to carry large objects	-	0.16 (2.24)	-	-
Driving style	0.30 (4.55)	0.23 (2.96)	0.25 (3.04)	0.25 (4.28)
Driver kindness	-	0.14 (1.82)	-	-
Hybrid buses	-	-	-	-
Noise	-	-	0.21 (2.24)	-
Mu(01)	1.73 (17.21)	1.52 (18.02)	2.31 (17.01)	1.73 (17.21)
Mu(02)	4.60 (33.49)	4.09 (37.05)	5.55 (30.2)	4.6 (33.37)
Log-likelihood	-494.86	-675.79	-603.9	-510.26
Psuedo-R ²	0.73	0.63	0.70	0.71
Degrees of freedom	17	19	20	15

Both the evaluation of general satisfaction (OS) and the satisfaction of the attributes has been measured following the same Likert scale. Therefore a comparison between the parameters of the same model can be made, understanding that a parameter of greater value will give greater importance to its corresponding attribute. The most influencing parameter is comfort (CM), which shows the highest parameter value for BASE, ATTRIBUTE and MI models. The comfort on board the bus is followed by the driving style (DS), which also represents how comfortable the ride is. Without considering the variables that turn out to be highly statistically insignificant, the ticket price (PR) show the lower parameter value, which means that the price to get to the service is not really important to define users' satisfaction. Service related attributes (travel time, waiting time...) show medium level values and those attributes that are clearly additional to the basic service, such as, special lines, cleanliness or noise turn out not statistically significant. The trend is similar in all models. Thereupon, it can be said that user satisfaction is highly defined by the comfort during the trip. Users could be used to the actual level of service and see it as acceptable. In such a way that users feeling more satisfied would come from attributes related with how comfortable the trip is.

When comparing different models, even taking data from similar sources and based on the same scale, because constant values and threshold parameters are different, a direct comparison is not possible. For this reason, all model parameters have been standardized before comparing them. In Figura 52 it can be seen that the correlation between the normalized models is considerable. In this case, without going into detail of each

parameter individually, it can be seen that the MI model shows a similar trend to the BASE model. That is, the normalized parameters vary jointly, obtaining high values in the MI model when the values are high in the BASE model and vice versa.

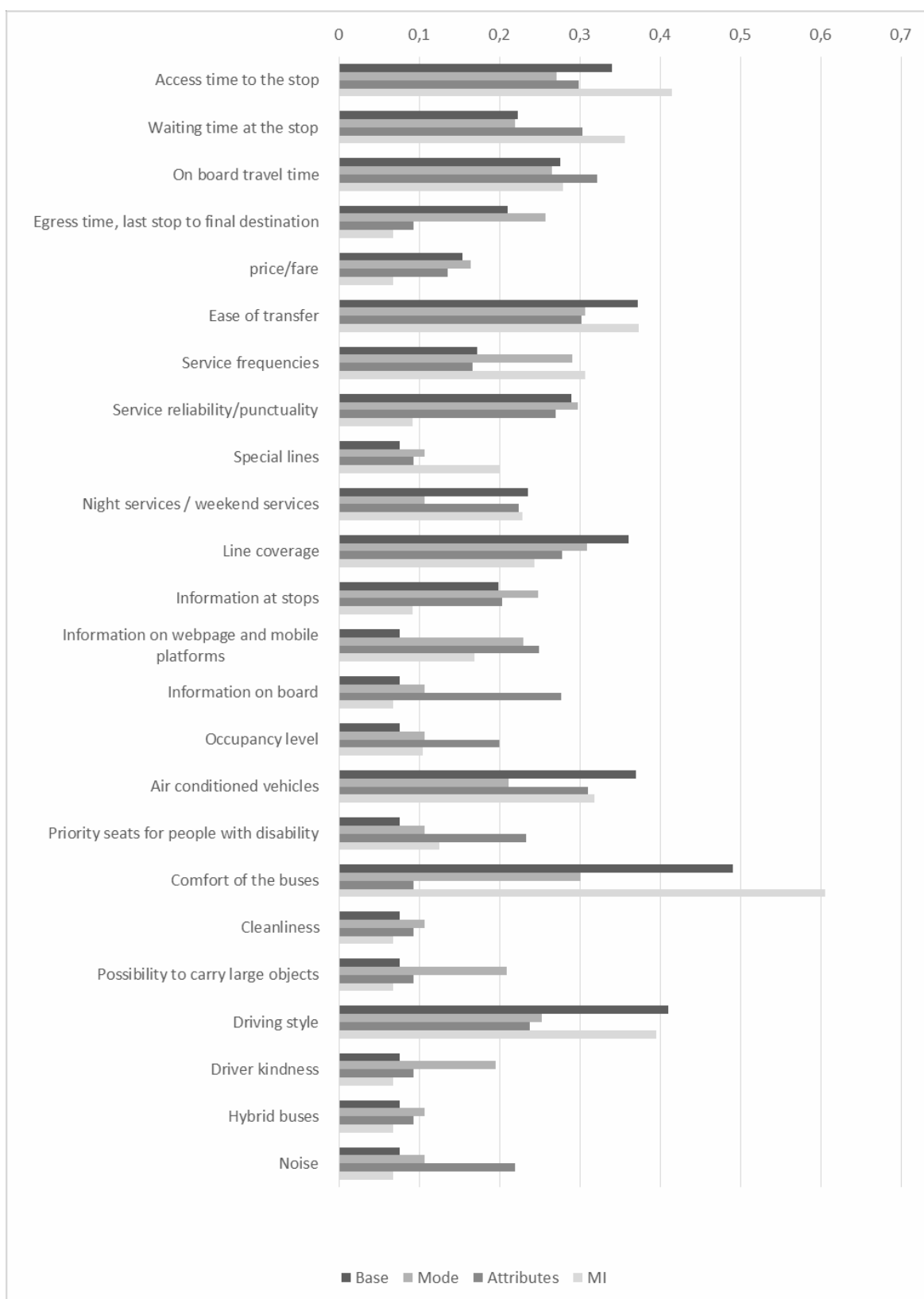


Figura 5 Artículo 2: Figure 2 - Normalized model parameters: a comparison of models with full vs. partial data

This is not the case for all the variables. There are some cases where the correlation between these models is not weak. For example: egress time (DT) and bus fare (PR) are not significant in the MI model but they are in the BASE model. The Pearson correlation coefficients calculated between models show a very high correlation, coefficient values compared to BASE model are 0.95 for MODE, 0.97 for ATTRIBUTE and 0.99 for MI.

Finally, it is essential to compare the explanatory ability that have been achieved with the different models based on goodness of fit indicators, such as Log-Likelihood value or count R^2 value (Echaniz et al., 2018; William H Greene and Hensher, 2010). Tabla 26 shows that the prediction capability of the BASE model is the best one, which shows a Log-likelihood value of -494.86 and count R^2 value of 0.73, that is, the model is able to explain 73% of the variation observed in the data. None of the models with missing data can match the prediction capacity that is obtained with the BASE model; however, the differences are relatively small, with the MI model having 1% less in predictive power. The MODE model has the worst predictive power, where the loss of accuracy is up to 10%.

An additional to predictive power, we carry a Vuong test (Vuong, 1989) to compare the models with results shown in Tabla 27. Z values close to 0 mean that the two models behave similarly. Absolute values greater than $Z = 1.96$ consider that the two models exhibit different behaviour at the 95% level of confidence. As can be seen, the BASE model based on full information outperforms alternative models fitted with partial information. The only model that behaves statistically similar to the Base model the MI model, with a z-value of 0.39. Therefore, MI model can be considered a slightly worse model than the BASE model but not different in a statistical sense. The other two models show values considerably larger than 1.96, and thus they are expected to behave differently from the Base model.

Tabla 27 Artículo 2: Table 4 - Vuong test for non-nested models

Z (Model 1 vs Model 2)		Model 2			
		Base	Mode	Attribute	MI
Model 1	Base	0.00	7.77	4.23	0.39
	Mode	-	0.00	-2.29	-7.21
	ATT	-	-	0.00	-3.83
	MI	-	-	-	0.00

5. CONCLUSIONS

This paper has shown a method to analyse public transport users' satisfaction based on partial information data. In addition, the empirical evidence included in this paper has shown that Ordered Probit Models, widely used in the analysis of users' satisfaction, can be estimated from a partial database with a minimum loss of information.

It has been observed that, even considering half of the available data, the descriptive analysis of attribute evaluations suffers a very small variation. Therefore, it can be said

that it is not necessary to collect all the data if what is wanted is to simply study the average satisfaction of the users. This is a common practice among public transport operators in order to obtain a clear picture of users' satisfaction towards their services. Thus, the lessons learned in this study could provide a considerable economic advantage for companies by requiring less time and resources to conduct the surveys. Having said that, it is necessary to emphasize that the results obtained in this study belong to a medium-sized coastal city, with a single public transport system (and operator) based on bus services, where the averages and deviations of the evaluations of the attributes and overall satisfaction of transport system are those that have been shown throughout the article. In consequence, caveats should be taken when extrapolating this study to cities or other public transport modes without prior analysis.

Regarding the modelling, the best methodology to fulfil the missing data turn out to be the Multiple Imputation (MI), which has allowed to obtain similar results to the ones obtained with the complete data. Vuong test carried out has shown that both models (the one obtained with the complete dataset and the one obtained after applying MI to impute the missing information) behave similarly.

The main output of this study is that the comparison between the models has shown that there is the possibility of obtaining very similar results with very similar fits to reality even starting from a partial information datasets. This allows to optimize the resources so that the time and the cost of the surveys can be reduced to a great extent, reducing the loss of information caused from the modelling of the data. Future studies will be focussed on reducing even more the need of data by applying different methodologies in order to obtain similar results as the ones obtained with ordered models.

Another way to optimize the surveying process is to study the possibility of applying different methodologies to obtain similar results obtained with the models used in this study, an issue that will be addressed in future studies.

ACKNOWLEDGEMENTS

This study has been possible thanks to the financing of the Spanish Ministry of Economy and Industry in the TRA2015-69903-R Project, the training grant FPU15 / 02990 of the Spanish Ministry of Education, Culture and Sports, the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 688082 - SETA Project) and thanks to the Spanish Ministry of Science, Innovation and Universities through the project TRA2017-85853-C2-1-R.

CONFLICT OF INTEREST

On behalf of all authors, the corresponding author states that there is no conflict of interest.

AUTHORS' CONTRIBUTION

E Echaniz: Modelling and manuscript writing C Ho: Manuscript writing and reviewing A Rodriguez: Data gathering and modelling L dell'Olio: Manuscript reviewing and editing.

4.3 ARTÍCULO 3: ADDRESSING THE IMPORTANCE OF SERVICE ATTRIBUTES IN RAILWAYS

4.3.1 Indicadores bibliométricos de la revista de publicación del artículo 3 en el JCR



sustainability

Título de la revista: Sustainability

ISSN: 2071-1050

eISSN: 2071-1050

Editorial: MDPI

País de publicación: Suiza

Volumen de publicación: 11(12)

Fecha de publicación: 21 de junio de 2019

Páginas: 3411

DOI: <https://doi.org/10.3390/su11123411>

Factor de impacto JCR

Tabla 28 Factores de impacto de la revista del artículo 3

Año de publicación (2018) ⁴	2,592
Promedio de 5 años	2,801

Ranking

Tabla 29 Ranking de la revista del artículo 3

Área de conocimiento	Posición	Percentil	Cuartil
Environmental Sciences	105/250	58,200	Q2
Green & sustainable science & technology	20/35	44,286	Q2

Numero de citas que posee el artículo

Tabla 30 Número de citas del artículo 3

Plataforma de medición	Número de citas
Google Scholar	2
Web of Science	1
Scopus	1

⁴ Se considerará como el valor del factor de impacto el establecido para el año de publicación del artículo. En el caso que el valor del año correspondiente no esté disponible se considerará el último año disponible.

4.3.2 Transcripción del artículo 3: Addressing the importance of service attributes in railways

ADDRESSING THE IMPORTANCE OF SERVICE ATTRIBUTES IN RAILWAYS

Roberto Sañudo, Eneko Echaniz, Borja Alonso and Rubén Cordera*

Department of Transportation, University of Cantabria, Av. de Los Castros 44, 39005 Santander, Cantabria, Spain; roberto.sanudo@unican.es (R.S.); eneko.echaniz@unican.es (E.E.); borja.alonso@unican.es (B.A.)

*Correspondence: ruben.cordera@unican.es;

ABSTRACT:

This research presents a combined qualitative and quantitative method focused on determining the level of importance that users place on different attributes of a railway service. The identification of these preferences will allow the establishment of policies that could increase the quality and demand of this mode. The qualitative analysis is based on a participation process through focus groups and in-depth interviews. From these sources, a diagnosis of the system and the main attributes that may influence the demand have been obtained. The quantitative part of the methodology is grounded on the realization of a stated preference survey using Best–Worst case 1 scaling. This method has made it possible to determine that the attributes related to the fare system, travel time, and intermodality are the most important to increase the quality of rail transport. Instead, the least important attributes were those related to very specific additional services. At the same time, it has been observed that the importance varies depending on the frequency of railway use. Thus, less frequent and nonusers considered the connections with other transport modes and user information more important. Therefore, improving intermodality could be a cost-effective way to attract new users.

Keywords: focus groups; railways; best–worst scaling; stated preference surveys; service attributes

1. INTRODUCTION AND OBJECTIVES

Current transport policies are focusing on promoting more efficient and environmentally friendly travel modes. The railway has proven to be a transportation alternative that meets this criterion. The roadmap of transport of the European Commission (2011) advocates for a promotion of this mode to the detriment of road transport, given its greater sustainability both environmentally and in terms of safety.

In order to encourage the use of the railway and to increase its future demand, it is necessary to establish which aspects are more appealing to the different types of transport users, i.e., which aspects are more important to determine service quality. In this research, a methodology is proposed to understand what are the key aspects of a railway service and how important are they for users.

Service quality studies have highlighted the importance of considering the perceptions of the users, trying to improve those aspects in which they are more interested in order to increase the demand of public transport (dell'Olio et al., 2017). In addition, the railway sector is now more open to competition both within the sector, given the liberalization measures promoted by the European Union (Bošković and Bugarinović, 2015; Kleinová, 2016), and with other transport modes (Albalade et al., 2015; D'Alfonso et al., 2015; Sun et al., 2017) which implies that it must pay more attention to users' satisfaction.

Several authors have developed service quality studies in the field of railways using satisfaction or Stated Preference (SP) surveys. Satisfaction surveys can contain questions about the perception or the importance of each service attribute as well as the overall service. These satisfaction surveys allow for users' opinions to be taken into account and avoids evaluating public transports services only from the point of view of the managers. The first satisfaction surveys were carried out in the air and urban transport sector (Andreassen, 1995; Ritchie et al., 1980) and have been much more rarely applied in the land interurban transport, including the railway sector. Givoni and Rietveld (2007) and Brons et al. (2009) performed a customer satisfaction survey highlighting the importance of considering the access and the quality of the facilities present in the stations to increase the demand of the railway mode. They argued that measures focused on improving these aspects can be more cost-effective than increasing the number or the quality of the railway services. In another case study, Nandan (2010) tried to collect evidence of the determinants of customer satisfaction with service quality of railway platforms at stations. The authors considered five determinants of customer satisfaction: information, behavioural factors, basic facilities, refreshments, and safety/security. The results achieved, using a sample of 700 railway users collected at 12 platforms of one station, showed how refreshments and behavioural factors (behaviour of railway staff) were the most important attributes for passengers. Considering the quality of the service, Nathanail (2008) studied the Hellenic Railways with the aim of measuring the level of quality of the different aspects and their importance to the users. Using 22 indicators, it was established that the service was performing high in safety, accuracy, and additional services but low in cleanliness and provision of information to users. In addition, the research found that the most important attributes for users were safety during the trip, itinerary accuracy, and train and station cleanness. In another case study in Italy, Eboli and Mazzulla (2015), using a satisfaction survey and a structural equation model, concluded that information, cleanliness, punctuality, and frequency of the services were the key aspects that should receive new investments according to the preferences of the users. Finally, Aydin et al. (2015) studied the situation of the rail transit systems of Istanbul through a customer satisfaction survey. Additionally, the authors conducted a questionnaire to professionals and academics in order to calculate the importance of the different criteria using a pairwise comparison. The results showed that safety, total journey time (waiting time, in-vehicle time, and itinerary accuracy) and accessibility were the most important criteria according to the experts.

SP surveys can also be applied to know the importance of different attributes in the quality of service (stated importance) in addition to its common application in explaining travel choices, making forecasts about the market share of different transport modes or estimating the value of time of users (Hensher, 2006; Román et al., 2007). In the field of

railways, Ahern and Tapley (2008) used a SP survey to simulate the choice between bus and train in intercity trips in Ireland. The authors concluded that the cost and travel time were the most important variables when explaining the modal choice. Park and Ha (2006) applied a similar methodology for intercity travel with competition between the high-speed train and the airplane using the fare, frequency, and access/destination time as attributes. The authors estimated that the discrete choice model based on SP data presented an error of 14% compared to the real demand of travellers who finally use the train and the airplane. Román et al. (2010) analysed the competition between the high-speed train and other modes available in the Madrid–Zaragoza–Barcelona (Spain) corridor using a mixture of SP data and revealed preference data (RP). The forecasts made by the authors were based on variables such as travel cost (fare), travel time, access time/destination time, frequency, comfort, and reliability. In this case, the demand was more responsive to travel time than to the fare or access/destination time. In Liu et al. (2016), the most influential variables were analysed when increasing the use of Maryland rail modes. These authors observed that the importance of the variables changed according to the type of transport analysed (metro, light rail, and commuter rail). For light rail stations, employment nearby, service level, feeder bus connectivity, stations located in the Central Business District (CBD), and terminal stations were significant factors affecting ridership, while for commuter rail, only feeder bus connections turned out to be significant. The studies presented above infer the relative importance of service attributes based on modelling results. However, the recent study developed by Cao and Cao (2017) shows that the importance levels can vary depending on if they are inferred from a modelling technique or directly asked to the users via a traditional rating.

The methodology proposed in this research is based on the combination of qualitative and quantitative techniques. Among the qualitative techniques, Focus Groups (FGs) and the in-depth interviews were chosen. Through this information, a diagnosis of the railway mode was made, obtaining a list of the variables that may influence the choice of railway by transport users. A similar procedure was applied in dell’Olio et al. (2013) to establish the variables that affect the behaviour of individuals in railway incidents and accidents. Among the quantitative methods, the use of an SP survey is proposed to elaborate a complete ranking of the importance of the variables related to the railway service, based on a Best–Worst (BW) case 1 scaling (Louviere et al., 2013).

The main contribution of this research is to develop and test a complete methodology for determining the most important attributes of the railway service to different types of transport users. In addition, it has been decided to use a BW survey technique (Louviere et al., 2013), which, as far as the authors’ knowledge, has not been directly applied before to the railway field or to define the importance of the different attributes belonging to the railway service. However, this method has been successfully used to analyse the importance of the different attributes belonging to a bus-based urban transport service. This was the case of the study made by Beck and Rose (2016), where the authors carried out a comparison between the satisfaction and the importance of the different attributes of the public transport service. These authors concluded that, in their case, the BW surveys allowed respondents to be more committed to conducting the questionnaire. In addition, it was possible to capture the difference given to different attributes between

importance and satisfaction and its correlation structure, unlike when using traditional rating methods.

In the following section, the proposed methodology is detailed, making use of both qualitative and quantitative techniques. In Section 3, this methodology is applied to a practical case study. The aim of this case study is to determine the importance of different attributes of the railway mode for users. This section will summarize the results obtained in the qualitative research, mainly through a SWOT (Strengths, Weaknesses, Opportunities, and Threats) analysis, as well as the results of the quantitative research based on the data collected by the BW survey. Finally, Section 4 summarizes the main conclusions derived from this study.

2. PROPOSED METHODOLOGY

The methodology applied in the study has been divided into two main phases (see Figura 6). The first phase, consultation or the qualitative phase, aims to obtain information from the agents involved on the railway system. The objective is to determine what are the main attributes that affect the choice of the railways compared to other modes. This first phase was based on the use of FGs and in-depth interview techniques. Different stakeholders were involved in this phase such as passengers, railway crew, and railway scholars/experts.

The second block of the methodology is based on the design and implementation of a SP survey based on a BW case 1 type responses. Using the data obtained from this survey, it is possible to order the attributes obtained in the qualitative phase according to their level of importance. These attributes were selected as key factors to enhance the quality and demand of the railway mode. In Figura 6, a complete scheme of the proposed methodology is presented.

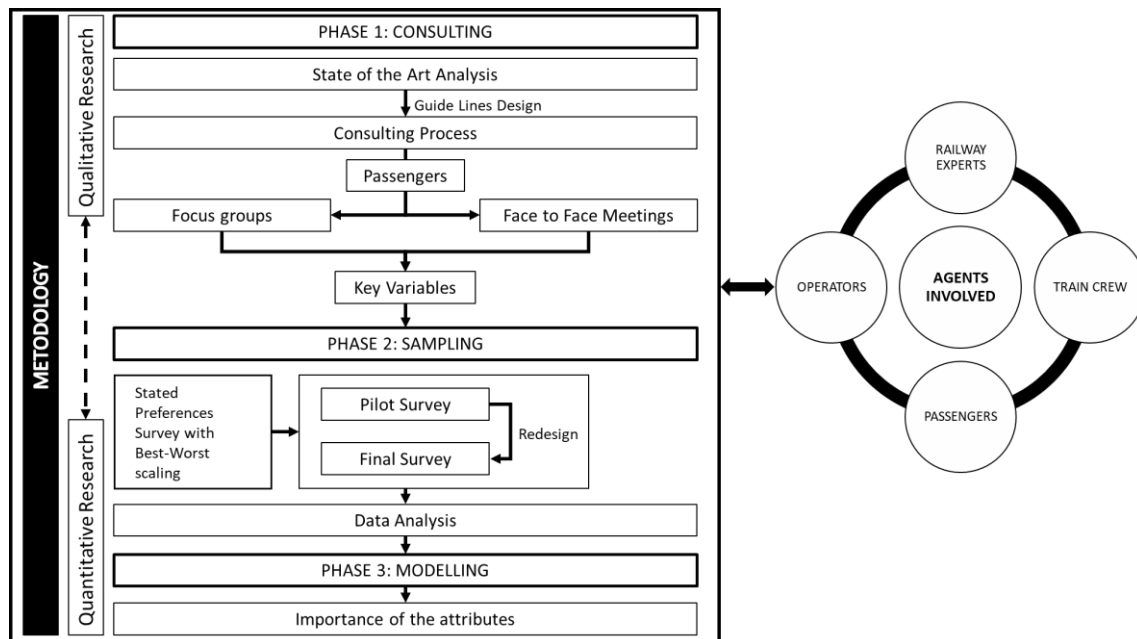


Figura 6 Artículo 3: Figure 1. Methodology proposed for studying the needs of transport users.

2.1. Qualitative Research

The qualitative research consists mainly of a consultation phase with the main agents involved in the railway sector. Prior to the consultation stage, a review of the existing literature was made. After that, the consultation process is developed by using two well-known techniques such as FGs and in-depth interviews.

A FG is a research method designed to explore a particular topic by gathering the experiences and perceptions of selected target populations (Ward and Atkins, 2002), where a moderator leads the group and guides the conversation (Bruseberg and McDonagh-Philp, 2002). Krueger and Casey (2000) define an FG as a carefully designed discussion providing people's perceptions on a particular area of interest. For example, Ibeas et al. (2011) used FGs and Mega Focus Groups (MFG) to find out people's opinions and perceptions on mobility within an urban environment. The methods for carrying out these in-depth interviews and FGs are fundamentally three (Rea and Parker, 2014): (i) Face to face, gathering groups of people to carry out FGs (of a minimum of 8 people) or through face-to-face interviews; (ii) by telephone or video-call; and (iii) via web through an online platform.

These three methods can be used to extract relevant information. However, each method has its own advantages and disadvantages that need to be considered. The face-to-face method is, among the three exposed, the one that allows gathering more direct information from the agents. Both the interviews and the FGs should be conducted by a skilled person in order to ensure that the quantity and quality of the information obtained is the highest possible. However, the realization of face-to-face interviews and FGs also implies additional costs derived from the need to travel and the use of facilities and additional material (recording tools and so on). The interview/FG format via telephone or video-call has its main advantage in allowing remote interaction with the respondent, although it is likely that it does not allow for the collection of as much information as the face-to-face method. Finally, online web formats have the advantage of being cheap to apply, although achieving the participation of agents through online platforms can be more complicated than through direct interview methods.

The choice of method for conducting the interviews and FGs therefore depends on the resources available and the advantages associated with each method. In general, the use of face-to-face methods is the most advisable to capture all possible information. In addition, it must be kept in mind that these techniques must be applied to all agents with significant relevance in the field of study. In the case of railway mobility, it would be advisable to have the information provided by passengers, experts (academics, operators, and entrepreneurs) and workers of the sector. Based on the above, the face-to-face interview format was chosen for this study.

The information collected through these qualitative techniques can be synthesized using different procedures. In this case, the application of the SWOT analysis is proposed (Hill and Westbrook, 1997) to make a general diagnosis of the railway sector. From the analysis of the consultation process, a selection of relevant attributes for the improvement of railway transport is obtained.

2.2. Quantitative Research

The quantitative research consists of the sampling and modelling phases (see Figura 6). Once the qualitative research is carried out, enough information is available to proceed with the design of the survey. This survey allows us to define what the most relevant attributes for users are. Therefore, the qualitative information is an essential requirement to carry out the qualitative analysis.

The survey serves to determine the order of importance of all the attributes indicated by the users. SP surveys are based on presenting hypothetical scenarios shown to different users in which they have to choose the best alternative according to their preferences (Rose and Bliemer, 2014). The use of the SP technique is necessary in this case, as traditional RP rating surveys are insufficient to capture the information and variability of the importance given to such a high number of attributes. SP surveys aim to capture the preferences of users regarding these attributes, being a quasi-experiment based on hypothetical situations established by the researcher and therefore being an approximation to a controlled experiment (Sanko, 2001).

The SP survey can be accompanied by a user characterization survey if it is required to differentiate between the preferences of different sociodemographic groups. Before carrying out the final characterization/SP survey, it is advisable to make a pilot survey to refine the questionnaire and to ensure the best design possible.

However, a large evidence shows that the assignment of attributes in an SP survey can have an impact on the responses and hence on the results of the models derived from them (Rose and Bliemer, 2014). Therefore, this assignment of attributes cannot be done randomly but must be applied using a systematic method that provides the best possible results according to the established objectives. Among the currently available methods, efficient techniques stand out which allow obtaining parameters in the models with the lowest possible standard error (Huber and Zwerina, 1996; Rose and Bliemer, 2009). Efficient designs are based on minimizing an efficiency indicator, the best known being the D-error:

$$D-error = \det(\Omega(X, \tilde{\beta}))^{1/k} \quad (50)$$

where Ω is the variance-covariance matrix, $\tilde{\beta}$ are prior parameters, and k is the number of parameters to be estimated in the model. This technique allows for the selection of, among all hypothetical scenarios, those that reproduce more robust parameters for the model that is intended to be estimated. The efficient design is useful to eliminate the problems related to classic factorial or orthogonal designs, presenting the advantage of reproducing robust models even when the respondents do not answer to all the scenarios presented (a relatively frequent case). To calculate the experimental design, specific software such as NGene can be used (Rose et al., 2009).

If the D-Error type of design is selected, the required sample is determined by the parameter of the model with the highest standard error:

$$N \geq \left(\frac{se(X, \tilde{\beta})t_{\alpha}}{\tilde{\beta}_k} \right)^2 \quad (51)$$

where N is the required sample, $se(X, \beta)$ is the standard error of a parameter, and t_α is the value of the T test for a given level of confidence.

The questionnaire should be designed to be answered by a sample of the target population, ensuring that the parameters obtained are representative. This type of surveys can be done through the following methods [2]: interception, telephone, online, or by ordinary mail. Each method has its advantages and disadvantages, although the online survey is the only feasible strategy with moderate costs in the case of a very broad geographical scope of people to be surveyed.

The data obtained from this SP survey can be later used in the analysis and modelling phase. In this case, a model based on BW (case 1) responses is proposed, where respondents have to indicate the most important and least important attributes for them from a predefined set of options. This type of choice allows to obtain more information than the traditional simple choice, and at the same time, it does not require as much cognitive load for the respondent as to make a complete ranking of all the attributes presented (Louviere et al., 2015). In this way, BW scaling can be considered a mid-point between the single choice and a complete ranking, balancing the main strengths and weaknesses of both kinds of responses.

The modelling of the data can be done following the Multinomial Logit (MNL) method, the most popular discrete choice model used in transport. In this model, random residuals are considered to be distributed according to a Gumbel distribution (Domencich and McFadden, 1975). The probability of choice is given by the following expression:

$$P_{BW}(i|i'|X) = \frac{\exp[u_i - u_{i'}]}{\sum_{j \neq i, j' \in K} \exp[u_j - u_{j'}]} \quad (52)$$

where the probability P of choosing the option i as best and the option $i' \neq i$ as worst from among a set X is based on the differences on utilities between all the alternatives $K = \{1, \dots, k\}$. These utilities u are defined generally by a vector of parameters to be estimated β . Thus, the parameter estimate β_k can be interpreted as the importance level of attribute k relative to the reference/base attribute which has $\beta_0 = 0$.

3. CASE STUDY AND RESULTS

3.1. Case Study Description

The methodology presented in the previous section has been applied to the case study of the NEAR2050 project (Shift2Rail Programme). The aim of the study has been to determine the importance that different attributes have for transport users when choosing the railway to make their trips. The action on the most relevant attributes should be preferential to increase the railway quality and demand. The project has focused on studying the European railway environment, considering all the distances covered by the railway transport (urban, intercity and long distance, understanding as urban railways metro, light rails, and tram services).

3.2. Qualitative Analysis Results

In the consultation phase of qualitative research, it was decided to carry out a total of 3 FGs and 7 in-depth interviews with various agents involved in the railway sector. The

interviews were conducted exclusively with experts (academics, managers, presidents of railway organizations, etc.). Given that the perceptions of users and the crew may be different, it was decided to carry out separate FGs, two with 8 users each and one more with 8 railway workers. Both in the interviews and in the FGs, those involved were asked to present their opinions considering a European scope. All the consultation phase was carried out from December 2016 to February 2017. From this qualitative research, two results were obtained:

- On the one hand, an analysis of Weaknesses, Threats, Strengths, and Opportunities of the railway sector (SWOT analysis).
- On the other hand, a list of attributes that can influence the choice of the railway mode by transport users. This list was also complemented with the results of the literature analysed.

3.2.1. SWOT Analysis

The results of the SWOT analysis conducted jointly from both interviews and FGs are shown in Tabla 31 (strengths and weaknesses internal to the railway sector) and Tabla 32 (opportunities and threats external to the railway sector). The SWOT analysis allows to obtain a preliminary diagnosis about the status of the railway sector according to the agents involved. The information shown in the SWOT represents the main aspect in which the different agents had a consensus about the different characteristics of the railway sector. This information was obtained through the analysis of the transcripts of the FGs and in-depth interviews.

Tabla 31 Artículo 3: Table 1. SWOT (Strengths, Weaknesses, Opportunities, and Threats) Analysis: Strengths and weaknesses extracted from focus groups (FGs) and in-depth interviews.

Strengths	Weaknesses
<ul style="list-style-type: none"> • Good information, accessibility, and services at the main stations both inside and outside the stations. • Adequate fares and fare information. With the possibility of purchasing tickets online and at the stations. • Good accessibility for People with Reduced Mobility (PRM) at the main stations. • Strategic location of some stations. • Greater comfort than other transport modes. • Good on-board services (cafeteria, Wi-Fi, music, video, silent cars, day care, etc.). • High levels of reliability in long- and medium-distance services. 	<ul style="list-style-type: none"> • Difficulty making transfers (especially for elderly people) due to a lack of information. • Short-distance intercity services have low frequencies and reliability with almost no additional services. • Additional cost for internet purchasing. Malfunction of the webpage. Inflexible rates. Scarce discounts for groups and families. Few reduced tickets. • Lack of services in non-main stations, especially for PRM. • Infrastructure problems with too long trains in non-main stations and different track widths. • Bad geographical connections. • Repetition of audio-visual content (movies, music, etc.). • Lack of comfort in some seats. • Restriction to travel with pets and large luggage, bicycles, etc. • Excessive baggage checks.

- High levels of security against thefts and incidents.
- High safety under unfavourable weather conditions.

Tabla 32 Artículo 3: Table 2. SWOT Analysis: Opportunities and threats extracted from FGs and in-depth interviews.

Opportunities	Threats
<ul style="list-style-type: none"> • Possibility of introducing low cost services. • Public financing of part of the management expenses to make prices more competitive compared to other modes of transport such as the airplane or the car. • Possibility of carrying out baggage checks only during times of high demand. • Improvement of ticket purchasing systems through new external technologies. • Possibility that local authorities improve accessibility to stations through public transport. • Promotion of future investments in double lanes for intercity trips. • Platforms technology to catch the train without waiting. • Reduction in the need for maintenance work at stations due to external automation. • Development of augmented reality services to provide better information at stations. • New prediction systems to develop on-demand express services and to react to delays or incidents. 	<ul style="list-style-type: none"> • Difficult to use websites. Automatic ticket selling systems are a barrier for some sectors of the population. • Location of stations outside the city. • Limited parking space at stations. Illegal parking in some stations. • Low frequency compared to the bus. • Lack of infrastructure maintenance derived from passivity in the improvement of services. • High cost of developing new infrastructure. • Non-human customer care and excessive automation. • Lack of confidence of the passenger due to poorly resolved incidents. • Badly planned transfers between countries. • New competitors in costs and safety for long-distance trips such as vactrain (vacuum tube train, i.e., Hyperloop project) and others.

Among the strengths of the system, aspects related to information, accessibility, or auxiliary services and comfort stand out, especially those related to the main station and services. Weaknesses, on the other hand, are more related to the tariff system and deficiencies in certain services, such as low frequencies, lack of services in non-main stations, and growing obsolescence in lines that are not adapted to high speed.

The opportunities mentioned in the rail system are the possibility of technological improvements in areas such as ticket acquisition systems, user access systems to trains, augmented reality services to improve information, or new systems that allow the prevention of delays or incidents. The possibility of introducing improvements to the fare system, such as new low-cost services, improvements in intermodality through parking and public transport, and more specialized on-board services, were also cited.

Within the threats are the competition with other current transport modes such as the bus, given its higher frequencies, and a possible future competition with new transport modes based on vacuum (i.e., Hyperloop) technology. The excessive use of technology was also mentioned as a threat in case of leading to an excessively impersonal customer treatment. Finally, the lack of interoperability between railway systems in different countries was also considered as a relevant threat.

3.2.2. Influential Attributes

A list of attributes was selected based on the interviews and FGs carried out. These attributes can affect, in a greater or a lesser extent, the choice of the railway mode when making a trip. These will be the attributes that will subsequently be ordered according to their importance by the users through the SP survey. Thirty attributes were defined and can be classified into the following groups (see the meaning of the acronyms in Tabla 33):

- Attributes related with the available information to passengers: IA, UT, IUT, OBI, LAN, and IAD.
- Attributes related with the total travel time of the trip: AT, TFT, TVT, NUM, ST, and LOC.
- Attributes related with the fare system and policy: PR and TICK.
- Attributes related with the comfort, security and additional services: COM, SERV, SEG, SOB, SPO, OBB, OBP, OBL, ACC, RE, and PUR.
- Attributes related with intermodality: CP, BP, CONU, and CONI.
- Attributes related with the environment: ENV.

Tabla 33 Artículo 3: Table 3. Attributes obtained from FGs and in-depth interviews.

Code	Acronym	Type	Attributes
1	IA	Information	Information to access the station
2	AT	Travel Time	Access time to get the station
3	TFT	Travel Time	Transfer time from one train to another
4	TVT	Travel Time	Travel time by train (duration of the trip)
5	COM	Comfort, services and security	On board comfort (comfortable seats, ergonomics, and on-board services)
6	CP	Intermodality	Car parking availability at the station
7	BP	Intermodality	Bicycle parking availability at the station
8	UT	Information	Information about urban transport connections at the station
9	IUT	Information	Information about intercity transport connections at the station
10	NUM	Travel Time	Number of daily and hourly trains
11	SERV	Comfort, services and security	On-board services (restaurant, sleeper cars, crèche, silent cars, Wi-Fi, plugs, etc.)
12	OBI	Information	On-board information (connections and transfers, stops, incidents, etc.)
13	ENV	Environment	Environmental policy of the company

14	PR	Fares	Fares (fare policy, discounts, low-cost services, etc.)
15	SEG	Comfort, services and security	On board service segregation (1st and 2nd class)
16	LAN	Information	Services in local language and in English (information and services available in multiple languages)
17	TICK	Fares	Ticketing system (electronic ticket, integrated ticket with other transport systems, etc.)
18	SOB	Comfort, services and security	Safety (on-board against accidents)
19	SPO	Comfort, services and security	Security (police controls at the stations, luggage control, etc.)
20	OBB	Comfort, services and security	Possibility of carrying bikes on board
21	OBP	Comfort, services and security	Possibility of carrying pets on board
22	OBL	Comfort, services and security	Possibility of carrying luggage with no weight limitation
23	ACC	Comfort, services and security	Accessibility for people with reduced mobility
24	RE	Comfort, services and security	Seat reservation
25	PUR	Comfort, services and security	Ease of ticket purchase (more sale points, vending machines, online purchase, etc.)
26	ST	Travel Time	Duration of the security controls when accessing the train (luggage control)
27	CONU	Intermodality	Connections with urban transport at the station (both for access and arrival)
28	CONI	Intermodality	Connections with interurban transport at the station (both for access and arrival)
29	LOC	Travel Time	Location of the station near city centre
30	IAD	Information	Information about activities at the destination (economy, leisure, services in destination, etc.)

3.3. Quantitative Analysis Results

Based on the information obtained in the qualitative research phase, in this section, we proceeded to design the SP survey. The later analysis and modelling of the gathered data allows to order the attributes according to their importance.

3.3.1. Survey Design

The final survey carried out by transport users consisted of four sections. In the first one, a general introduction was presented, explaining the main objectives of the study. In section two, the respondents were asked about their sociodemographic characteristics (gender, age, city of residence, possession of driving license, if they suffer from reduced mobility or not, work status, and income level). In the third section, information was asked about railway use: frequency of use, distance to the nearest station, modes used to get to and from the stations, mode of transport most commonly used, and in case of not using the train, the reason not to do so.

The last section of the survey corresponds to the SP scenarios. Given that a complete ranking of the 30 attributes obtained in qualitative research would require a lot of cognitive effort for the respondents, the SP methodology allowed them to compare only 4 attributes each time. In each SP scenario, the respondent decided which one of the four attributes was the most important attribute (Best option) and which one was the least important (Worst option). This kind of choice is commonly known as BW. The efficient experimental design obtained after carrying out the pilot survey with the minimum D-Error can be consulted in Tabla 34. The design was divided into 6 blocks; that is, each respondent had to answer only one block consisting of 5 BW scenarios, choosing through 4 attributes in each.

Tabla 34 Artículo 3: Table 4. Experimental design of the definitive Stated Preference (SP) survey (the codes of the attributes can be consulted in Table 3).

Block	Scenario	Attribute 1	Attribute 2	Attribute 3	Attribute 4
1	1	20	28	13	2
1	2	11	19	8	24
1	3	14	11	3	18
1	4	3	7	23	5
1	5	19	25	26	15
2	1	17	13	15	14
2	2	30	17	5	4
2	3	6	2	14	10
2	4	9	21	10	26
2	5	26	8	16	20
3	1	5	6	24	25
3	2	13	3	1	12
3	3	24	12	29	21
3	4	28	23	6	22
3	5	7	24	22	27

4	1	18	20	25	9
4	2	12	9	27	30
4	3	27	26	28	3
4	4	16	10	7	29
4	5	2	27	19	17
5	1	23	16	12	19
5	2	15	30	18	7
5	3	22	15	21	1
5	4	8	1	30	6
5	5	29	18	17	8
6	1	10	5	11	13
6	2	25	29	4	28
6	3	21	14	20	23
6	4	4	22	9	11
6	5	1	4	2	16

The survey was conducted through an online platform in order to obtain greater heterogeneity in the responses and to have a much wider sample.

Scenario 3

★ From the following railway characteristics, choose the one you consider is the most important and the one you consider least important.

	Most Important	Least Important
Car parking availability at the station	<input type="radio"/>	<input type="radio"/>
Travel time to get to the station	<input type="radio"/>	<input type="radio"/>
Fares (Fare policy, discounts, low cost services...)	<input type="radio"/>	<input type="radio"/>
Number of daily and hourly trains	<input type="radio"/>	<input type="radio"/>

Figura 7 Artículo 3: Figure 2 shows an example of a BW choice scenario between four attributes.

3.3.2. Sociodemographic Information of Respondents

The final survey was carried out in May 2017, obtaining 316 complete answers from a total of 609 participants, that is, a 52% of complete answers from European countries. The survey was developed on an online platform and disseminated via different social networks, where people all over Europe were invited to participate. No incentive was

offered to participate in the survey. Responses from over 24 countries were obtained related to more than 180 cities. The respondents turned out to be mostly men (79%) who reside in medium-sized cities (30%), of young age, with a majority between 25 and 34 years. Almost all respondents had a driver's license, but not all (70% of those with a license) had their own vehicle. In addition, respondents were mainly employees (55%) with an average salary. Regarding the use of the railway, a great number of the respondents used the railway at least a few times a month and most of them had a station near to their home or place of residence. The complete breakdown of the data is shown in Tabla 35. Despite the sample not being representative of the European population, the observations were weighted considering sex and age strata in order to assure its validity. In this way, the sample was adapted to represent the population structure of the European Union: 49% males and 51% females; 26% of the population of less than 25 years, 42% between 25 and 54 years, 13% between 55 and 64 years, and 19% older than 65 years (Central Intelligence Agency, 2017).

Tabla 35 Artículo 3: Table 5. Sociodemographic characteristics of respondents.

Gender	Male	79%
	Female	21%
City Size	Rural areas (pop < 25,000)	17%
	Small cities (pop 25,000–150,000)	20%
	Medium size cities (pop 150,000–500,000)	30%
	Large cities (pop 500,000–1,500,000)	17%
	Metropolis (pop > 1,500,000)	16%
Age	<25	6%
	25–29	21%
	30–34	15%
	35–39	13%
	40–44	11%
	45–49	11%
	50–54	7%
	55–59	6%
	60–65	4%
	>65	6%
Driving License	Driving Licence	96%
	Of which own a car	70%
Suffer from reduced mobility or has a disability		3%
Work Status	Employee	55%
	Self-employed	9%
	Public employee	17%
	Student	12%
	Retired/Pensioner	5%
	Unemployed	2%
	Other	0%
Income Level	<900 €	10%
	900–1500 €	10%
	1500–2500 €	26%
	2500–3500 €	13%
	3500–4500 €	6%

	>4500 €	10%
	No answer	25%
Train Usage	Never/occasionally (0–6 times per year)	38%
	Weekly (1–2 times per week)	31%
	Frequently/daily (Every day or almost every day)	31%
Distance to the Nearest Station	<1 km (5–10 min walking)	30%
	1–3 km (10–30 min walking)	39%
	3–6 km (30–60 min walking)	19%
	>6 km (more than 1 h walking)	12%

3.3.3. Modelling

With the BW data collected, an MNL model has been specified, which allows the establishment of the order of importance of the attributes based on the coefficient estimated. The higher the coefficient of an attribute, the greater the importance that respondents have given to it. For the model to be estimable, it is necessary to establish an attribute as a base, to which the coefficient value of 0 is given. In this case, the attribute OBP (Possibility of carrying pets on board) was chosen given that it was considered the least important. Hence, the coefficients of all the remaining attributes will have positive values. Tabla 36 shows the parameters estimated.

Tabla 36 Artículo 3: Table 6. Multinomial Logit (MNL) model derived from the Best–Worst exercise.

Code	Attribute	Coefficient	z Value
1	IA	1.639	8.33
2	AT	2.859	13.41
3	TFT	3.135	15.28
4	TVT	4.216	18.8
5	COM	3.244	15.2
6	CP	1.814	8.85
7	BP	1.672	8.17
8	UT	2.850	14.16
9	IUT	2.406	11.53
10	NUM	4.039	18.7
11	SERV	2.629	12.38
12	OBI	3.068	15.13
13	ENV	1.439	6.88
14	PR	4.076	18.55
15	SEG	0.294	1.58
16	LAN	1.690	8.25

17	TICK	2.645	12.5
18	SOB	2.516	12.42
19	SPO	2.030	10.03
20	OBB	1.703	8.46
21	OBL	1.705	8.63
22	ACC	2.179	10.71
23	RE	1.803	9.27
24	PUR	3.010	14.59
25	ST	1.997	9.96
26	CONU	3.405	16.54
27	CONI	3.817	18.14
28	LOC	3.359	16.25
29	IAD	0.266	1.22
Log-likelihood		-3096.43	
AIC/N		3.956	

The difference between the values of the parameters is associated with how much an attribute is more important compared to another. To make it easier to understand, the coefficients have been normalized with values between 0 and 100, giving the value 100 to the highest coefficient (TVT) and 0 to the lowest (OBP). Figura 8 shows the result of that normalization, ordering the attributes from the most important to the least important.

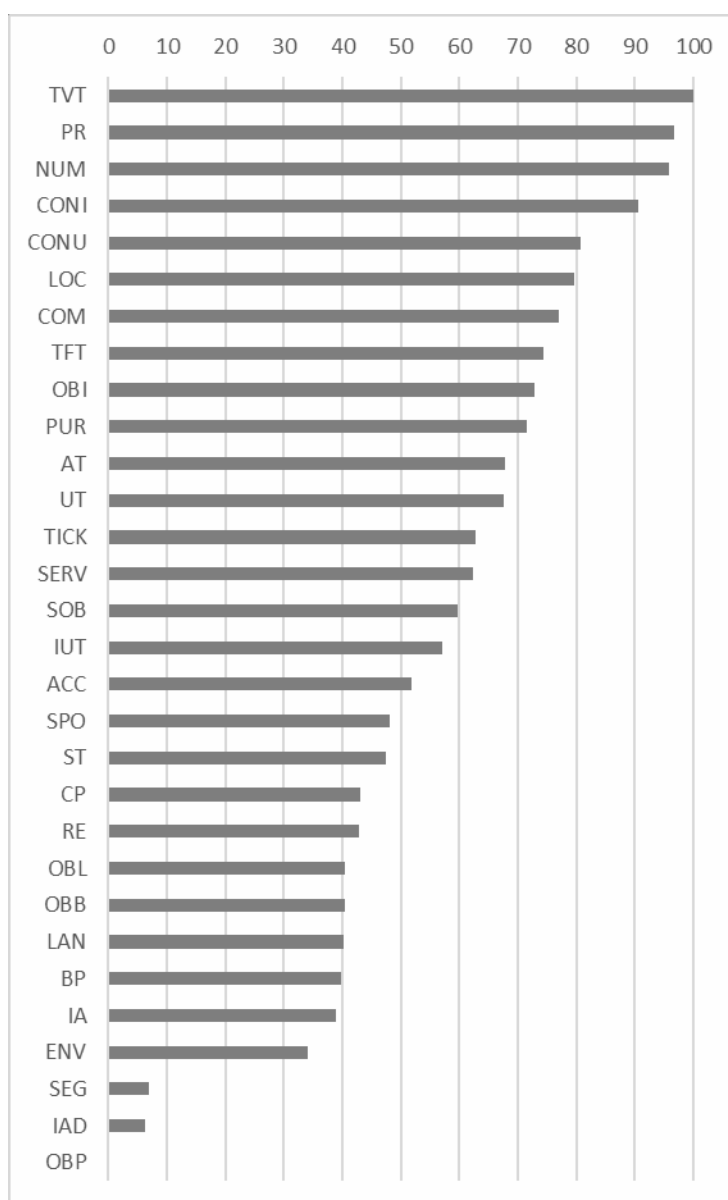


Figura 8 Artículo 3: Figure 3. Importance of the attributes according to their normalized coefficient.

The most important attribute turned out to be the travel time (TVT), followed closely by the price (PR) and frequency of services (NUM). In a second level are the attributes associated with accessibility to railway service, considering both interurban and urban connections of the stations (CONI and CONU, respectively) and the good location of these (LOC), often near city centres.

The least important attributes turned out to be those attributes that are very specific and purely additional services, such as the segregation of the service in different classes (SEG), the information about activities at the destination (IAD), and the possibility of taking pets on board (OBP). Furthermore, it can be seen that there is a considerable difference between these three less important attributes and the group of immediately important attributes, where the respondents grant a level of relevance more or less similar to a set of 7 attributes related to additional rail services: Seat reservation (RE), possibility of carrying luggage with no weight limitation (OBL), possibility of carrying bikes on board (OBB), information and services available in multiple languages (LAN), bicycle

parking availability at the station (BP), information to access the station (IA), and the environmental policy of the company (ENV).

Another way to analyse the results is to perform the Best minus Worst operation. That is, subtract the number of times an attribute has been chosen as the least important option to the number of times it has been chosen as the most important one. In Figura 9, it can be seen that there is a clear correlation between the Best minus Worst operation and the parameters obtained by the MNL Model with an $R^2 = 0.94$. This method was used to analyse the attributes in a grouped form, according to the classification defined in Tabla 33.

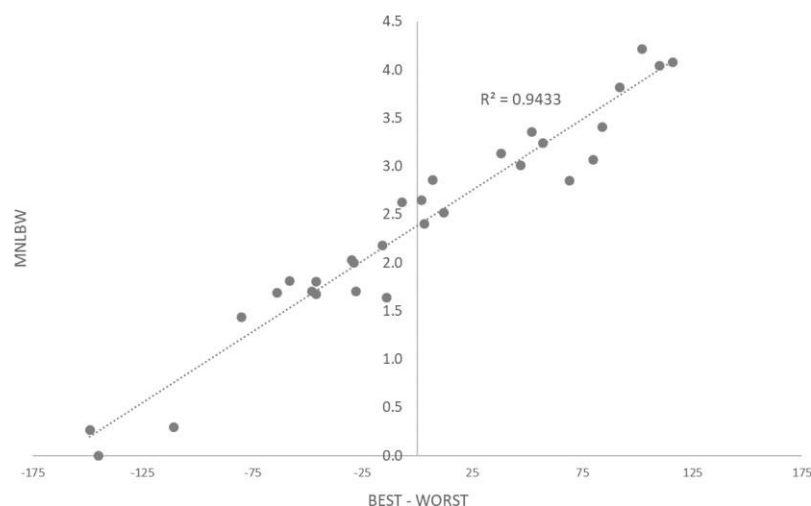


Figura 9 Artículo 3: Figure 4. Correlation between the parameters obtained in the MNL model and the Best–Worst operation.

Figura 10 shows the level of importance given to each one of the six groups. The groups are defined by three data: (i) the maximum Best minus Worst value (best rated attribute within the group), (ii) the minimum Best minus Worst values (worst rated attribute), and (iii) the average value of Best minus Worst of all the attributes within each group. A group that has a high Best minus Worst average shows that the attributes that are part of that group are very important for users. In contrast, the maximum and minimum values give a view of the dispersion on the importance of the attributes within each group.

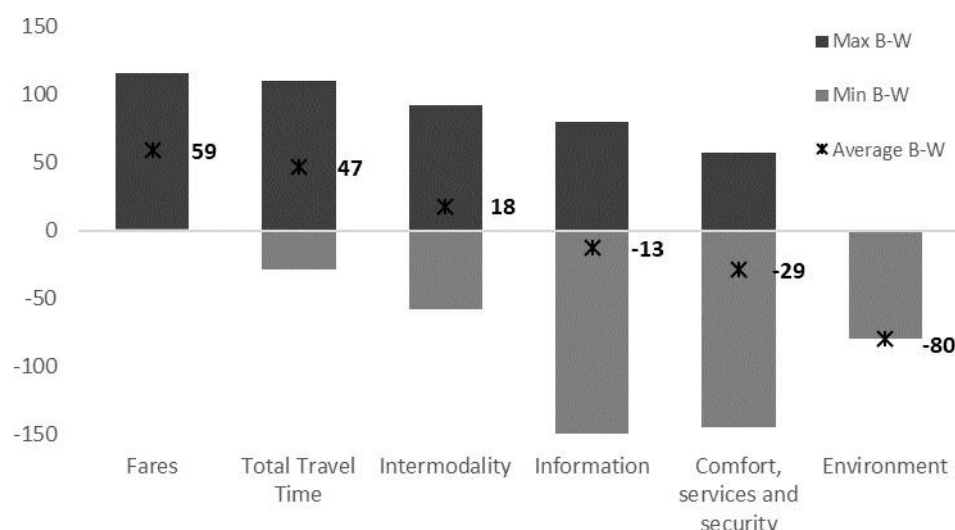


Figura 10 Artículo 3: Figure 5. Average Best–Worst values and maximum and minimum given by the respondents by type of attribute.

The fare and the fare policy in general turned out to be the most important group of attributes. However, although the price of the trip is the most important attribute, a competitive price must be accompanied by a good ticket purchasing system with the possibility of using different types of tickets such as physical, electronic, and preferably combined with other modes of transport.

The attributes belonging to total travel time also showed to be very important for users. In addition, even though the level of importance changes, all the times associated with a trip are important for the user. Therefore, any effort to avoid wasted time along the full journey will have a very positive effect on the perception of the transport users. The only time attribute that was not important for the respondents was the duration of the security controls, probably because the time used for these purposes increases the security sensation of the users that see the controls as necessary.

Intermodality was considered important when combining with other public transport systems, while car parking at stations are not really important for users. Having parking space for bicycles is less important than being able to take the bicycle on the train, so it can be said that the people who make a trip combining the bicycle and railway prefer to travel taking the bike with them inside the train, instead of, for example, taking a public bicycle for the first and last miles.

In the case of information, the respondents consider information regarding the connections of the stations with other public transport systems and information on-board important, while the possibility of having information in several languages and information about the activities available at the destination received very low scores.

The comfort and auxiliary services are generally considered quite insignificant. The only attributes associated with this group that are considered important are on-board comfort and on-board safety against accidents. The rest, although different in level, are initially not significantly important for users.

Finally, the environmental policy is not considered too important for the overall population, although certain segments of the population can show greater interest in this aspect.

Therefore, it can be said that the desired service for transport users consists of a railway system with competitive prices and travel times. In addition, the railway must be comfortable and well-connected with other transport modes, especially with public ones, and users must have all this information easily accessible and clearly explained in all possible points and sources (stations, on-board, mobile, etc.). Once those needs are met, investing efforts can focus on other types of additional services that complement the passenger's trip.

The socioeconomic variables obtained in the survey have allowed an analysis considering different types of user to be carried out. The most interesting results are those related with railway usage levels since they allow the differentiation of the importance given to the attributes by railway users and nonusers. For modelling, the different usage levels have been coded using the effect coding technique. Unlike dummy coding, in this type of coding, the reference group is identified by a -1 value. Therefore, the constant term obtained in the model represents the grand mean of the importance of each attribute for the sample as a whole. Thus, the coefficients associated with each level show the deviation of that group from the mean. The advantage of this type of coding is that it is possible to calculate the coefficient associated with the reference group. The value of this coefficient is calculated as the sum of the negative values of the coefficients of the rest of the groups. The effect coded variables have been introduced in the model as interactions with each attribute. Occasional/nonuser group have been defined as the base group. Tabla 37 shows the estimated model, where those interactions that have turned out not to be statistically significant have been omitted.

Tabla 37 Artículo 3: Table 7. MNL model with interactions by railway usage level.

Attribute	Constant	z Value	Frequent/Daily	z Value	Weekly	z Value	Never/Occasionally
IA	1.664	8.40	-	-	-	-	-
AT	2.908	13.49	-	-	0.466	2.81	-0.466
TFT	3.176	15.38	-	-	-	-	-
TVT	4.307	18.86	-	-	0.363	1.99	-0.363
COM	3.282	15.28	-	-	-	-	-
CP	1.777	8.54	-0.567	-3.66	-	-	0.567
BP	1.723	8.36	-	-	0.427	2.78	-0.427
UT	2.893	14.27	-	-	-	-	-
IUT	2.449	11.64	-	-	-	-	-
NUM	4.150	18.82	0.595	3.24	-	-	-0.595
SERV	2.653	12.42	-	-	-	-	-
OBI	3.127	15.28	-	-	-	-	-
ENV	1.490	7.07	-	-	0.282	1.63	-0.282
PR	4.135	18.52	-0.275	-1.49	-	-	0.275
SEG	0.320	1.71	-	-	0.339	2.16	-0.339
LAN	1.736	8.42	-	-	-	-	-
TICK	2.682	12.58	-	-	-	-	-
SOB	2.557	12.53	-	-	-	-	-

SPO	2.052	10.06	-0.409	-2.64	-	-	0.409
OBB	1.755	8.69	-	-	0.417	2.50	-0.417
OBL	1.772	8.82	-	-	0.364	2.34	-0.364
ACC	2.216	10.78	-	-	-	-	-
RE	1.833	9.33	-0.359	-2.02	0.245	1.62	0.113
PUR	3.051	14.69	-	-	-	-	-
ST	2.036	10.08	-	-	-	-	-
CONU	3.450	16.63	-	-	-	-	-
CONI	3.870	18.25	-	-	-	-	-
LOC	3.488	16.55	0.430	2.66	-	-	-0.430
IAD	0.291	1.33	-	-	-	-	-
Log-Likelihood						-3057.83	
AIC/N						3.925	

The model coefficients show the differences of each group for every particular attribute. In order to facilitate the visualization, the following graphs (Figura 11, Figura 12 and Figura 13) show the comparisons of each group with respect to the model estimated with the complete sample (Tabla 36). The graphs are sorted from highest to lowest frequency of railway use.

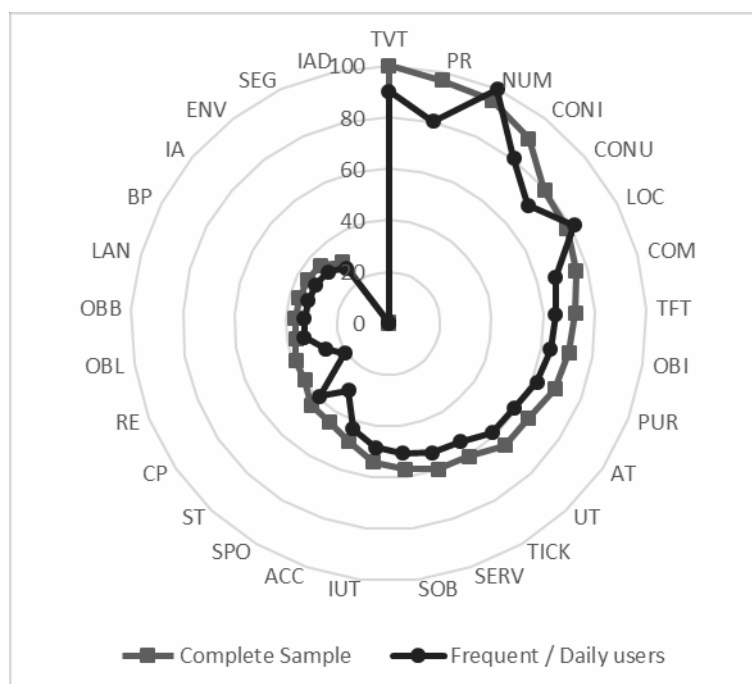


Figura 11 Artículo 3: Figure 6. Comparison of the difference in importance given to the attributes by the complete sample and frequent users.

The main difference between frequent users and the rest (Figura 11) is the clear reduction in the importance of ticket prices (attribute PR). In the same way, the travel time slightly reduces its importance, while the service frequency becomes the most important variable. In terms of accessibility and connections with other means of transport, whether urban or interurban, the importance is slightly more reduce, while the location of stations becomes more important. At the bottom of the list, there is a clear reduction in the importance associated with police security, the availability of car parking at stations, and the reservation of seats. Through these variations, we can deduce that regular rail users are those that use the railway for daily commuting. Therefore, the importance of the attributes

focuses mainly on having a good frequency of services and that the stations are accessible to their origins and/or destinations.

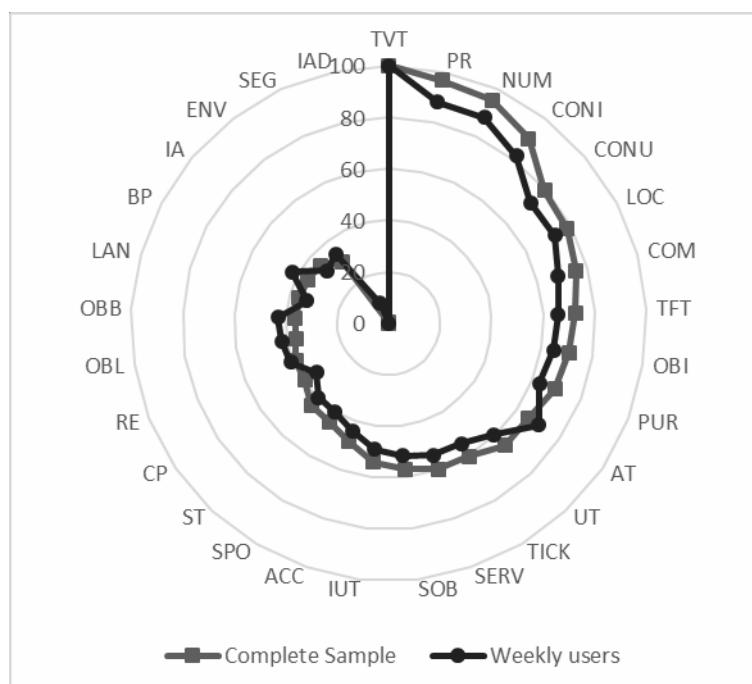


Figura 12 Artículo 3: Figure 7. Comparison of the difference in importance given to the attributes by the complete sample and weekly users.

Users who use the rail weekly (Figura 12) show a clear preference for total journey time, i.e., travel time by train and access time to it. In relation to additional services, a greater importance is observed for services related to intermodality with cycling and the possibility of carrying large volume luggage on board. We can include in this group those people who make their daily trips usually by bicycle but who combine it with the railway for different reasons such as a longer travel distance. On the other hand, this group can also include those people who make medium long-distance journeys and who choose the railway if it is competitive with the other modes in terms of travel and access time.

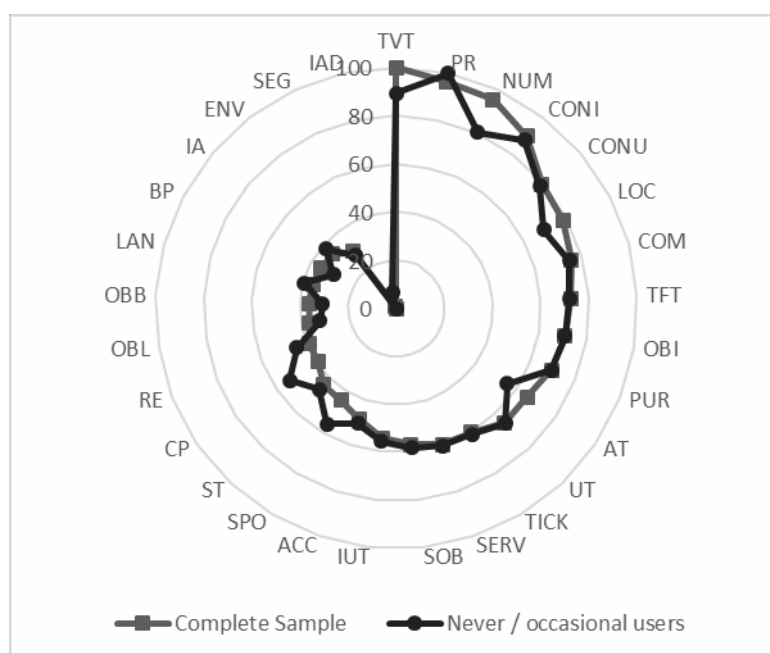


Figura 13 Artículo 3: Figure 8. Comparison of the difference in importance given to the attributes by the complete sample and occasional users/nonusers.

Finally, for those people who never or hardly ever use the railway (Figura 13), the price takes the greatest importance. In addition, the frequency, the location of the stations, and the access time considerably reduce their importance. In contrast, urban and interurban connections with other means of transport and the availability of car parking at the stations (CP) increase their importance. Additional services focused on unusual users are also considered more important, such as the availability of information in different languages, information to access stations, or information of activities at the destination. Within this group, we can include passengers who use the railway if it is economically better than the rest of the alternatives. In addition, these users appreciate the additional services that provide access to the necessary information of the service.

4. CONCLUSIONS

In this study, a methodology based on both qualitative and quantitative techniques has been applied in order to obtain which attributes are most relevant for transport users when choosing the railway mode. This information can be useful to increase the demand of this transport mode in the future compared to other less sustainable alternatives.

By using FGs and in-depth interviews, it has been possible to see how the perception of rail transport is quite different depending on the type of service. Users and experts generally perceive that high-speed services at main stations are reliable and have sufficient frequencies, auxiliary services, and information. However, in local services or in non-main stations, frequency and auxiliary services are perceived as deficient. The remarkable automation of the service, which could increase in the coming years, generates a certain level of distrust in some users when perceiving the customer care as too impersonal. This could become a barrier in increasing demand in the future. The current fare system is perceived as too rigid and not very sensitive to the differences between the types of users and trips, without the possibility to make transfers with other companies or trips in an easy and connected way. An improvement in fare flexibility and

giving the possibility of buying tickets to cover a complete trip (door-to-door and mobility as a service paradigm) and in combination with low-cost services could therefore attract more users.

When ordering attributes after the application of the Best–Worst scaling, it can be concluded that the total travel time is a key aspect for choosing rail transport. In second place are the price or fare policy and the frequency of the services, which show a similar importance and clearly, due to their great relevance, must accompany a good travel time. The accessibility to the railway service shows a great significance, focusing especially on the intermodality that the stations have with other public urban and interurban services, besides having a good location within the cities. Auxiliary services during the trip show very varied levels of importance, with, for example, the information on board or the comfort of the seats being important aspects, while the information regarding the destination, the possibility of acquiring tickets in different categories, or the possibility of travelling with pets on board show much lower levels of importance. The results obtained show that to improve the quality of rail transport, it is necessary to invest mainly in the infrastructure/services and to act on pricing, leaving aside all those auxiliary services that are very appreciated by some users but do not fulfil the need of the general users.

The analysis according to frequency of railway use show some differences in the importance given to the attributes. Frequent users, i.e., those who use it for daily commuting, do not consider price as such an important variable, while frequency of services and location of stations gains much more importance. Users who use it sporadically consider travel time as a key variable, this one being the most important attribute with a considerable difference from the rest. In the same way, access time to stations also shows a greater importance than for the rest of users. As for less frequent users and nonusers, price becomes the most important variable as well as urban and intercity connections with the stations and the availability of car parking in them. In addition, nonusers considered additional services (access information, information on activities at the destination, and information in different languages) to be more important than average.

To apply policies focused on less frequent and nonusers could be the easiest way to increase railway usage in Europe. In addition, these policies will highly unlikely discourage current passengers to continue using railways. In this way, the most cost-effective way to attract new users would be to improve intermodality, both with public and private modes, and information about the railway services. This result supports the conclusions obtained by Givoni and Rietveld (2007) and Brons et al. (2009) who argued that better access to the stations, intermodality, and information could attract new users.

To sum up, the combination of qualitative and quantitative methods has made it possible to emphasize the importance of focusing on aspects such as travel time, the flexibilization of the fare system, and the improvement of intermodality, especially to attract new users. Although the experience carried out in this study has focused on establishing the importance of attributes for the railway field, the methodology used can be applied to other public and private transport systems.

AUTHOR CONTRIBUTIONS

Conceptualization, R.S., E.E., and R.C.; methodology, E.E. and R.C.; validation, R.S., B.A., and R.C.; formal analysis, E.E.; investigation, R.S., E.E., and R.C.; resources, B.A.; data curation, R.S. and E.E.; writing—original draft preparation, E.E. and R.C.; writing—review and editing, R.S. and B.A.; visualization, E.E. and R.C.; supervision, B.A.; project administration, R.S.; funding acquisition, R.S. and B.A.

FUNDING

This research was supported by the Shift2Rail initiative through the financing of the NEAR2050 project (Project Reference: 730838, Call Reference S2R-OC-CCA-01-2015); the Spanish Ministry of Economy and Industry through the Project TRA2015-69903-R (co-funded by ERDF funds); and the Spanish Ministry of Education, Culture, and Sports through the training grant FPU15/02990.

ACKNOWLEDGMENTS

The authors would like to show their gratitude to Angel Ibeas of the University of Cantabria for his support during this research.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.



© 2019 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

4.4 ARTÍCULO 4: COMPARING BEST-WORST AND ORDERED LOGIT APPROACHES FOR USER SATISFACTION IN TRANSIT SERVICES

4.4.1 Indicadores bibliométricos de la revista de publicación del artículo 4 en el JCR



Título de la revista: Transportation Research Part A: Policy and Practice

ISSN: 0965-8564

Editorial: Pergamon-Elsevier SCI LTD

País de publicación: Reino Unido

Volumen de publicación: 130

Fecha de publicación: diciembre de 2019

Páginas: 752-769

DOI: <https://doi.org/10.1016/j.tranpol.2017.05.006>

Factor de impacto JCR

Tabla 38 Factores de impacto de la revista del artículo 4

Año de publicación (2018) ⁵	3,693
Promedio de 5 años	4,371

Ranking

Tabla 39 Ranking de la revista del artículo 4

Área de conocimiento	Posición	Percentil	Cuartil
Economics	34/363	90,771	Q1
Transportation	6/36	84,722	Q1

Numero de citas que posee el artículo

Tabla 40 Número de citas del artículo 4

Plataforma de medición	Número de citas
Google Scholar	0
Web of Science	0
Scopus	0

⁵ Se considerará como el valor del factor de impacto el establecido para el año de publicación del artículo. En el caso que el valor del año correspondiente no esté disponible se considerará el último año disponible.

4.4.2 Transcripción del artículo 4: Comparing best-worst and ordered logit approaches for user satisfaction in transit services

COMPARING BEST-WORST AND ORDERED LOGIT APPROACHES FOR USER SATISFACTION IN TRANSIT SERVICES

Eneko Echaniz

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

email: echanize@unican.es

(Corresponding author)

Chinh Ho

The University of Sydney, Institute of Transport and Logistics Studies, The University of Sydney Business School, 378 Abercrombie St, Darlington NSW 2008 (Australia)

email: quoc.ho@sydney.edu.au

Andres Rodriguez

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

email: andres.rodriguez@unican.es

Luigi dell'Olio

University of Cantabria, Transport Systems Research Group, E.T.S. Ingenieros de Caminos, Canales y Puertos, Avda. de los castros S/N 39005 Santander, Cantabria (Spain)

email: luigi.dellolio@unican.es

ABSTRACT

Customer overall satisfaction towards a public transport system depends mainly on two factors: how satisfied they are with different aspects that make up the service and how important each of the service aspects is to the customer. Traditionally, researchers use revealed preference surveys and ordered probit/logit models to estimate the contribution of each service attribute towards the overall satisfaction. This paper aims to verify the possibility of replacing the traditional method with the more cost-effective best-worst case 1 method, using a customer survey recently conducted in Santander, Spain. The results show that the satisfaction level obtained from these alternative methods are remarkably similar. The relative importance of each attribute delivered by the two methods differ, with the Best-Worst approach showing more intuitive and consistent results with the literature on public transport customer satisfaction. A regression method

is developed to derive customer satisfaction with each service attribute from Best-Worst modelling results.

Highlights:

- Conventional satisfaction ratings and Best-Worst results are strongly correlated.
- Estimated attribute importance levels vary across the methods used.
- Best-Worst importance results are more in line with the existing literature.
- Satisfaction rating results can be inferred from Best-Worst data.
- Best-Worst model delivers better and more consistent results for decision making.

Keywords: User satisfaction; Importance; Public transport; Best-Worst scaling; Ordered Logit.

1. INTRODUCTION

Ordered logit and probit models have been widely used in public transport satisfaction studies (Alonso et al., 2018; Bordagaray et al., 2014; Echaniz et al., 2018). These models predict the overall quality of a transport service (i.e., overall satisfaction) based on the extent to which users are satisfied with each of the service attributes such as travel time, spatial coverage or service frequency. Thus, satisfaction data on each of the attributes that define the service have to be collected, usually based on a Likert scale on which respondents indicate their level of satisfaction with each service attribute and overall. Such surveys are usually lengthy and repetitive, resulting in low response rate or loss of sample due to respondent burden. For example, in a recent on-board surveys of bus users in Santander, Spain (Echaniz et al., 2018) many respondents were not able to finish the questionnaire based on traditional rating method. Consequently, interviewers had to leave the bus with the respondent in order to complete the survey. Not only reducing the interviewer's productivity but also complicating the logistics of survey.

An alternative approach to studying customer satisfaction and the contributions of different service attributes to the customer satisfaction is the Best-Worst (BW) case 1 method. Rather than asking the respondent to evaluate each service attribute at a time, as in the traditional rating method, the BW method shows the respondent a set of service attributes at the time and asks them to choose the best and the worst of the attributes shown and the process is repeated until all attributes are covered. Thus, BW surveys are less time consuming and more intuitive for respondents, requiring less guidance from the interviewer. Hence, BW survey method is a cost-efficient way to obtain data required for satisfaction studies, and yet this method has not been very used to study customer satisfaction in public transport systems, with only few exceptions (Beck et al., 2017; Beck and Rose, 2016).

This paper investigates the possibility of replacing the traditional approach to customer satisfaction with a best-worst case 1 survey method (Louviere et al., 2015). If this potential of BW method is verified, a more flexible and effective way to study customer satisfaction will be achieved, without imposing much burden on the respondent (through shorter questionnaires) or compromising the statistical robustness of the results. With respect to the latter, pioneer work of Beck and Rose (2016) suggests that the relative

importance of service attributes identified by conventional rating methods can be biased. This limitation of the traditional method must be overcome, in Cao and Cao (2017) for example it was found that implicit importance (attribute importance derived from modelling estimates) based on ordered logit models is presumably a better indicator for decision making than the traditional rating. Motivated by these observations, this paper aims to compare the suitability of BW methods with that of traditional rating methods in studying customer satisfaction within the context of bus services.

The remainder of this paper is structured in to five sections. The next section reviews the relevant literature. Section 3 describes the data, the survey design and the sample. Section 4 explains the models, with estimation results presented in section 5. The paper ends with the conclusions of the main findings, with some directions for future study.

2. LITERATURE REVIEW

Budget constraints mean that investment options to improve services should be examined to maximise potential benefit for the given budget. One way to identify the most potential areas for improvement is to establish the extent to which users are satisfied with different service attributes. This can be achieved by asking users to give their level of satisfaction towards each service attribute through a traditional satisfaction survey with some ranking scale such as a 5-point Likert. However, investments should not be made based solely on a low level of satisfaction towards a certain attribute, since not all aspects of the service affect the users in the same way, with some service attributes being more important than others. A common solution in public transport (PT) customer satisfaction study is to ask the users how satisfied they are with the service alongside with how important each service attribute is to them. In doing so, the operator knows where to focus the improvement efforts to get the biggest benefit.

Customer satisfaction in PT services has traditionally been studied using revealed preference (RP) surveys. Different inception and interview methods have been used, such as on-board interviews (de Oña et al., 2013; Eboli and Mazzula, 2009; Eboli and Mazzulla, 2011; Echaniz et al., 2018), online surveys (Abenoza et al., 2017; Beck and Rose, 2016; Rose and Hensher, 2018) and mobile app (Guirao et al., 2015). Users evaluate different aspects of the PT service based on their experience using it, typically referred to as service attributes, based on qualitative or quantitative scales. These evaluations define the level of satisfaction users have with the different aspects of the PT service and also with the service bundle as a whole. The attributes to be evaluated by the respondent are usually derived from previous studies, as in the case of Hensher et al. (2003) where 13 representative items of the service were defined, or in Efthymiou et al. (2018) where factors that affect satisfaction are analysed for times of economic crisis. The subsequent analysis of the obtained data has been carried out in different ways over the years. The simplest and most popular method is descriptive analysis in which the mean and deviation of the level of satisfaction with each attribute and the entire service (i.e., overall satisfaction) are used to represent customer satisfaction. Thus, most studies on user satisfaction perform this basic analysis before applying a more complex and advanced method (e.g., de Oña et al., 2017; Eboli et al., 2018; Eboli and Mazzulla, 2015; Gonzalo-Orden et al., 2011; Tyrinopoulos and Antoniou, 2008).

The importance users place to different PT service attributes can be established using different methods, but they could be grouped into two approaches. The first approach uses the stated importance directly (explicit importance), where the customers rate each attribute on an importance scale similar to the one used for rating satisfaction. The second approach use derived importance in which the relative importance of each attribute is inferred by statistically analysing the relationship of individual attributes with the overall satisfaction (implicit importance) (Weinstein, 2000). Numerous studies were found to use the first approach to obtain the importance of the attributes (Beck and Rose, 2016; Guirao et al., 2016; Rose and Hensher, 2018) although this approach leads to an increase in the duration of the questioning, which means that surveys can become excessively long.

However, there are more efficient ways to obtain the importance level of the attributes, as is the case with the Best-Worst (BW) scaling (Louviere et al., 2015). Three types of BW exercises are used in the literature. These are the object case (Case 1), the profile case (Case 2), and the multi-profile case (Case 3). In Case 1, the respondent is asked to select the best and worst options from a series of objects or items such as a list of brand names, without showing any attribute or characteristic apart from the item itself. In Case 2, the respondent is asked to select the best and worst from a list of attributes, each of which was assigned a specific level determined by some experimental design. The choice is made between the different attributes with each one having their own set of levels. Case 3 is associated with the classic discrete choice experiments, where the choices are made between a set of alternatives composed of different attributes with different levels. The choice is then made between the different alternatives available which are composed by the same attributes but with different levels on each one. BW method have been successfully used in different transport studies. For instance, Beck et al. (2017) used a BW case 3 study to identify the respondent's attitudes towards choosing electric vehicles in the presence of regular fuelled alternatives. Or in Mulley et al., (2014) where BW case 1 was used to study the preference of the citizens regarding the construction of a Bus Rapid Transit (BRT) or a Light Rail Trains (LRT) service. In the context of public transport service satisfaction, Beck and Rose (2016) applied also BW case 1 method and compared it with the traditional rating method. They concluded that the traditional way of establishing the importance of the variables is biased in which the service attributes people most satisfied with were associated with the highest levels of importance. In turn, it was observed that traditional responses did not show great variability. Beck and Rose (2016) also concluded that the traditional rating did not provide enough information for decision making. On the other hand, the BW method made it possible to better capture these variations. The correlation between satisfaction and importance was found to be much more coherent using the BW method, proving to be a more useful decision tool. Following Beck and Rose (2016) this study uses a BW case 1 method instead of a traditional method.

The analysis of the BW data has been made following two main techniques. In one hand, score measures i.e. analytical closed form (ABW) (Lipovetsky and Conklin, 2014) or normalized BW scores (NBW) (Louviere et al., 2015). On the other hand, discrete choice models have been applied, such is the case on Beck and Rose (2016) or Marley and Pihlens (2012). Most used discrete choice model is the Multinomial Logit. In Marley et al. (2016) an interesting comparison is made between ABW and NBW score measures,

also, they compare the results obtained with a MNL model. The results show that for all three methods ABW, NBW and MNL are very close related.

Regarding the implicit importance, over the years different modelling methods have been used in order to find the most correct way to define the importance of the different attributes that define a service. de Oña and de Oña (2015) and dell’Olio et al. (2018) provide a comprehensive review of the methods. Recently, Allen et al., (2018b) and Allen et al., (2019) uses the structural equations model to study user satisfaction with Transantiago and Metro Madrid respectively. Mouwen (2015) performed an analysis of the public transport satisfaction in the Netherlands, the method used was the multiple regression. Discrete choice models have been also used, specially ordered models, both Logit and Probit (Allen et al., 2018a; Bordagaray et al., 2014; dell’Olio et al., 2010; Echaniz et al., 2018). Another way based on discrete choice modelling has been by carrying out stated preference (SP) surveys, which show the respondent a number of choice tasks and ask them to choose the one they most preferred. The data is analysed using discrete choice models of some kind. For instance, Román et al. (2014) used multinomial logit model (MNL) and Mixed Logit (ML) model to examine public transport services in Gran Canaria (Spain). Similarly, dell’Olio et al. (2011) used an MNL model to analyse the quality desired by future users.

Regardless of the data analysing method, the most statistically significant factors in most of these studies are the frequency of the service, the reliability and travel time, and to a lesser extent, the comfort of the buses and the smoothness of the ride. Allen et al. (2018) showed that having a reliable service and a good frequency were the most influential attributes when explaining the users’ satisfaction with the public bus system. The perceived waiting and travel times were also found to be of great importance. In addition, Mouwen (2015) showed that on-time performance, travel speed and frequency are most important attributes when explaining the overall quality of the service. Similarly, Román et al. (2014) showed that urban users has a greater willingness to pay for waiting time, travel time and access time. In dell’Olio et al. (2011) it was shown that in order to attract users it is necessary to increase the overall quality of the system by improving the waiting time, the comfort during the trip, the sources of information and the frequencies.

Thus, it appears that the majority of studies on public transport service satisfaction arrive at similar conclusions regarding the main drivers of customer satisfaction, even when using different modelling methods and datasets. Only one study was found that compares the results of an Ordered Logit model with a conventional rating based importance level (Cao and Cao, 2017). The main finding of that study was that the importance levels obtained with both methods were different. To the best of the authors' knowledge, no study has analysed the relationship between conventional rating and modelling with BW data focused in public transport satisfaction.

3. DATA

3.1. Survey

The survey includes two parts. The first part asked respondents to report their socio-economic characteristics (e.g., gender, age, work status, income level), level of bus usage (e.g., trip purpose and number of trips per week) as well as the availability of alternative modes for these trips. The second part involved the respondent evaluating the importance

and satisfaction towards different service attributes. A total of 24 service attributes, shown in Tabla 42, were used to define the services based on the existing international literature and several focus groups carried out in the city of Santander (Ibeas et al., 2011). These service attributes were grouped into six sets of four attributes each. Each respondent was asked to evaluate three sets of attributes allocated dynamically and randomly such that no attribute appeared twice for the same respondent and the total sample provides a balance of attributes assessed.

Each respondent was asked to evaluate the same set of attributes based on both traditional ranking (5-point Likert scale) and best-worst response mechanisms. In the traditional rating exercise, the respondent was asked to rate each attribute on a 5-point Likert scale (3 sets \times 4 attributes/set = 12 attributes in total). In the best-worst scaling exercise, the respondent was asked to select, out of the same four attributes included in the choice task, which attribute they are most and least satisfied with (satisfaction choice), as well as which attributes are most and least important to them (importance choice). Figura 14 summarise the data collected from one of the three choice tasks showed to each respondent. Respondents were not shown the three tasks of one question simultaneously but one after the other. At the end of the survey, all respondents were asked to rate the service as a whole, defined as Overall Satisfaction. The overall satisfaction was obtained following the same 5-point rating scale used for the attributes.

Ejemplo	Escala varloración (escala Likert de 5 puntos)						Elección Best-Worst			
	Muy Mal	Mal	Normal	Bien	Muy Bien	N/A	Más Importante	Menos Importante	Más Satisfactorio	Menos Satisfactorio
Atributo 1					X		X		X	
Atributo 2			X							
Atributo 3				X				X		
Atributo 4			X							X

Figura 14 Artículo 4: Figure 1 – Example data collected from one choice task based on Rating vs. BW scales

3.2. Sample

The surveys were run between October and November 2017 in the city of Santander. Face-to-face interviews were conducted on four bus lines operated by the municipal public company in the urban area of Santander. A total sample includes 808 completed interviews, spreading across the whole day with interviews taken place both at bus stops and on board.

Tabla 41 shows the main characteristics of the respondents. Women are over-represented in the sample (two in three respondents). A quarter of respondents are under 25 years old, while other age groups are more balance. Regarding occupation, almost half (47%) of the respondents were employed and nearly a quarter (24%) were students with the balance being retirees (17%), unemployed (8%) and housewives (5%). Half of the respondents had some other motorized alternative to make the same journey, while only a 6% would be willing to make the same journey by bicycle. The trips captured in the survey showed the important role of bus services in Santander for commuting purposes, i.e. travelling between home and office/school. Nearly half of the trips (46%) have the home as an origin and more than a quarter (29%) have the home as a destination. Work was the second main

reason, both as an origin and a destination. The respondents are mainly habitual users with a low frequency of use per day. Specifically, more than half (54%) of the respondents use bus services up to 15 time per week. Finally, due to the sensitivity of the question, 42% of respondents decided not to answer the question related to their income level. Of the people who did answer, income levels have a good mix, with a greater proportion of people with an average salary: 20% of people between 900 and 1500 € per month and 17% between 1500 and 2500€.

Tabla 41 Artículo 4: Table 1 - Respondents' socio-economic information

Gender	Male	33%	
	Female	67%	
Age	< 25	25%	
	25 - 34	14%	
	35 - 44	15%	
	45 - 54	17%	
	55 - 64	15%	
	65 - 75	11%	
	> 75	4%	
Work status	Housekeeper	5%	
	Employee	47%	
	Unemployed	8%	
	Student	24%	
	Retired	17%	
Other available transport systems	Car (Driving)	35%	
	Car (accompanying)	12%	
	Bike	6%	
	Motorcycle	3%	
	Other	44%	
Trip purpose		Origin:	Destination:
	Home	46%	29%
	Work	22%	25%
	Studies	9%	13%
	Health	4%	5%
	Shopping	5%	7%
	Leisure	10%	13%
	Other	5%	9%
Number of trips made by bus per week	< 5	26%	
	5 - 15	54%	
	15 - 30	18%	
	> 30	1%	
Monthly income	< 900€	7%	
	900€ - 1500€	20%	
	1500€ - 2500€	17%	
	> 2500€	14%	
	No answer	42%	

3.3. Rating scale results

Tabla 42 shows the user satisfaction with each of the 24 service attributes in a descending order. The traditional rating scale is based on a 5-point Likert scale and the rating is recoded to have the value from 0 (very unsatisfied) to 4 (very satisfied) for descriptive and modelling analysis.

Tabla 42 Artículo 4: Table 2 – Satisfaction ratings

Order	Attribute	Acronym	Mean	Mode	Std. Deviation
1	Use of hybrid buses	HY	3.24	3	0.77
2	Access time to bus stop	AT	2.94	3	0.90
3	Egress time from alighting stop to destination	DT	2.91	3	0.89
4	Vehicle cleanliness	CL	2.81	3	0.70
5	Ease of transfer	TR	2.79	3	0.90
6	Information at stops	IS	2.76	3	0.97
7	Information on board	IB	2.73	3	0.90
8	Comfort of the buses	CM	2.71	3	0.74
9	Service reliability / punctuality	SR	2.70	3	0.86
10	Driver's kindness	DK	2.63	3	0.85
11	Quality of bus stops	ST	2.62	3	0.81
12	Information on mobile app	IM	2.61	3	1.27
13	Line coverage	LC	2.60	3	0.83
14	Information on the web page	IW	2.58	3	0.96
15	Priority seats for people with reduced mobility (PRM)	RM	2.51	3	0.89
16	Waiting time	WT	2.50	3	0.91
17	Crowding level	OC	2.50	3	0.87
18	Readability of map design	MD	2.48	3	0.98
19	In-vehicle travel time	TT	2.47	3	0.85
20	Service frequency and timetables	SE	2.44	3	0.97
21	Driving style	DS	2.39	3	0.86
22	Price / Fare	PR	2.33	3	0.94
23	Calefaction / air conditioning	CA	2.31	3	0.99
24	Noise	NO	2.28	2	0.82
	Overall satisfaction	OS	2.69	3	0.80

Overall, the respondents are quite satisfied with the service, with an average rating of 2.69 out of 4. The service attribute that users are most satisfied with relates to the use of sustainable propulsion engines with hybrid vehicles (HY), being the only attribute that has the level of satisfaction exceeding 3. In total, nine attributes have a level of satisfaction greater than average, with the lowest level of satisfaction observed for noise level (NO), air conditioning (CA) and fare (PR).

4. MODEL SPECIFICATIONS

This section describes the modelling approaches used to model customer satisfaction data obtained from the empirical survey. First, an Ordered Logit model is discussed and shown how it is used for modelling the data obtained from the traditional rating responses. This is followed by a specification of two standard logit models for the empirical data obtained from BW responses: one model for the level of satisfaction and another model for level of importance.

4.1. Ordered Logit for traditional raking data

Ordered Logit models are based on the following specification of a latent regression:

$$q_i^* = \beta' \mathbf{x}_i + \varepsilon_i, i = 1, \dots, n. \quad (53)$$

in which the latent continuous preference variable q_i^* is only observed in discrete form q_i through a censoring mechanism:

$$\begin{aligned} q_i &= 0 \text{ if } q_i^* \leq \mu_0 \\ q_i &= 1 \text{ if } \mu_0 < q_i^* \leq \mu_1 \\ &\dots \\ q_i &= J \text{ if } \mu_{J-1} < q_i^* \leq \mu_J \end{aligned} \quad (54)$$

Note that the specification in equations (1) and (2) assumes that neither parameters β nor thresholds μ vary across individuals. This assumption of homoscedasticity is arguably strong and can be relaxed. The vector \mathbf{x}_i is a set of K covariates that are assumed to be independent of ε_i ; and β is a vector of K parameters to be estimated, together with $J + 2$ threshold parameters μ_j using N observations. The assumption of the disturbance ε_i completes the model specification. The conventional assumptions are that ε_i is continuous, random and follows a certain cumulative distribution function (CDF), $F(\varepsilon_i|\mathbf{x}_i) = F(\varepsilon_i)$.

For this study, q_i^* represents the non-observable overall satisfaction of the PT service, while q_i is the observable overall satisfaction obtained from the traditional rating question asked at the end of the survey. J represents the 5-point Likert scale options of the rating task shown in Figura 14; \mathbf{x}_i are the satisfaction ratings of the service attributes assessed by the respondent i with values ranging between 0 (“Very Bad”) and 4 (“Very Good”).

The dependent variable of the model is defined as the overall satisfaction (OS), while the independent variables are the satisfaction levels of the different attributes of the service. In total 24 independent variables or service attributes have been defined; however, conducting on-board face to face surveys means that a large proportion of bus users would not have enough time to provide their level of satisfaction towards each of the 24 attributes. Thus, each respondent was asked to rate 12 of the 24 attributes, which generates an additional modelling challenge but that can be solved using imputation methods to complete the database. The method used to complete the sample has been based on Multiple Imputation (Rubin, 1978, 1977), explained below. Echaniz et al. (2019) have shown that it is possible to obtain Ordered satisfaction models by using this method and that results obtained using this method are similar to those obtained with a complete database.

Multiple imputation is estimated by using a procedure called Fully Conditional Specification (FCS), which uses an iterative Monte Carlo method with Markov chains (van Buuren, 2007). The FCS approach is based on variable-by-variable imputation of data, specifying an estimation model for each one of the variables with missing data. The FCS tries to define $P(X, C, R|\theta)$ by specifying a conditional density $P(X_i|C, X_{-i}, R, \theta_i)$ for each X_i , this density is used to impute X_i^{mis} given some C, X_{-i} and R . An imputation consists of a complete cycle through all X_i (van Buuren, 2007). Where X represents the evaluation of the attributes, C the characterization variables, θ the parameters of the imputation model and R an indicator that show if X is a missing or observed value. The imputation is made by using the Gibbs sampling methodology (Casella et al., 2016; Gilks et al., 1996) assuming that the conditional density distribution exists. This methodology has been used in a large number of simulation studies (Brand, 1999; Horton et al., 2016; Raghunathan et al., 2001; Van Buuren et al., 2006) that have provided sufficient evidence that the results obtained through the FCS are generally unbiased and have adequate coverage.

Once the database is completed, the models are estimated as usual. To estimate the model some normalizations are required. First, to keep the positive sign of the probabilities it is required to $\mu_{j+1} > \mu_j$. As the variable q_i^* exists in the entire real line and the model contain a constant term $\beta_0 = 0$, it is necessary to define $\mu_0 = 0$ and $\mu_J = +\infty$. As. The data does not contain information about the scaling of the dependant variable q_i^* , therefore, the free variance parameter $Var(\varepsilon_i) = \sigma_\varepsilon^2$ cannot be estimated. The usual approach is to assume that σ_ε is constant and depends on the distribution assumed for ε_i . In Logit models it is assumed that ε_i follows a logistic distribution, resulting in $Var(\varepsilon_i) = \pi^2/3$. The associated probabilities are defined as:

$$Prob[q_i = j | x_i] = Prob[\varepsilon_i \leq \mu_j - \beta' x_i] - Prob[\mu_{j-1} - \beta' x_i] > 0, j = 0, 1, \dots, J \quad (55)$$

The model (3) is estimated using maximum likelihood estimator which maximises the log-likelihood function defined as follows:

$$\log L = \sum_{i=0}^N \sum_{j=0}^J m_{ij} \log \left[F(\mu_j - \beta'_k x_{ki}) - F(\mu_{j-1} - \beta'_k x_{ki}) \right] \quad (56)$$

where $F(\cdot)$ is the cumulative distribution function; $m_{ij} = 1$ if $q_i = j$ and 0 otherwise.

4.2. Multinomial Logit for Best-Worst scaling

The literature defines three type of BW data, known as Case 1, Case 2 and Case 3 as reviewed in section 2. Aiming to verify the importance of different service attributes associated with bus services without including different attribute levels, this paper adopts the BW Case 1 method.

There are a total K attributes to be chosen on the survey. In each BW task a subset Y of four attributes is shown. With the answers of the choice task, a vector $\delta = (\delta_1, \dots, \delta_k)$ is estimated, which is the utility coefficient of each attribute.

The probability of choosing an attribute $b \in Y$ as best is denoted as $P_B(b | Y)$. In the same way, the probability of choosing an attribute $w \in Y$ as worst is denoted as $P_W(w | Y)$. The *joint* probability of choosing attribute b as best and attribute $w \neq b$ as worst is defined as

$P_{BW}(bw | Y)$. In the experiment, the respondent had to select the best option and the worst option out of a subset of service attributes. The survey instrument was programmed in such a way that the respondent cannot advance if the same attribute was selected as both best and worst attributes. That is, the probability of the same attribute - call it x - being chosen as best *and* worst by the same respondent i (suppressed from the notations for simplicity), is always 0. Mathematically, $P_{BW}(xx|Y) = P_B(x|Y)P_W(x|Y) = 0$ since either $P_B(x|Y)$ or $P_W(x|Y)$ or both must be 0.

Adopting a standard logit specification to describe the choice of the best and the worst attributes, (i.e. assuming that the unobserved components of the utility follow Type 1 Generalized Extreme Value or Gumbel distribution with random variables independently and identically distributed), the probability $P_{BW}(bw | Y)$ for one BW choice task is defined as (5), which also is called the Maxdiff model (Marley and Louviere, 2005):

$$P_{BW}(bw|Y) = \frac{\exp[v(b) - v(w)]}{\sum_{l \neq k} \exp[v(l) - v(k)]} \quad (57)$$

where $v(\cdot)$ is the observable utility components specified as a linear-in-parameter function of attributes such as $v(k) = \delta_k y_k$ where y_k is an indicator vector of 0 and 1 ($y_k = 1$ when the attribute k is shown to the respondent i and 0 otherwise). In this way, the parameter estimate δ_k could be interpreted as the importance or satisfaction level (depending on which model is being analysed) of attribute k relative to the reference/base attribute which has $\delta_0 = 0$.

The maxdiff model assumes that the respondent simultaneously chooses the best and the worst options; however, it may be possible that the respondent selects the best option first, then eliminate this attribute out of the choice set before selecting the worst option. In this case, the repeated best-worst model specification as in equation (6) may be more appropriate (Dyachenko et al., 2012).

$$P_{BW}(bw|Y) = P_B(b|Y)P_W(w|Y - \{b\}) = \frac{\exp v(b)}{\sum_{l \in Y} \exp v(l)} \cdot \frac{\exp -v(b)}{\sum_{k \in Y - \{b\}} \exp v(k)} \quad (58)$$

Similarly, if we assume that the respondent selects the worst option first, then the best option, a repeated worst-best model as in equation (7) could be used (Dyachenko et al., 2012).

$$P_{WB}(wb|Y) = P_W(w|Y)P_B(b|Y - \{w\}) = \frac{\exp v(w)}{\sum_{l \in Y} \exp v(l)} \cdot \frac{\exp -v(w)}{\sum_{k \in Y - \{w\}} \exp v(k)} \quad (59)$$

Equations (58, 59) are alternative model specifications of equation (57). Empirical study (see for example Greene, 2016; Ho and Hensher, 2017) shows that these alternative model specifications are likely to produce similar results. Thus, this paper uses the Maxdiff model and acknowledges that the repeated best-worst or repeated worst-best specifications could be used as an alternative specification.

5. RESULTS

5.1. Discrete choice models

Three models were estimated using Nlogit v6.0: two MNL models based on the BW data obtained from the BW choice exercises and one OL. Tabla 43 shows the parameter estimates with t-values shown in parentheses. The BW models show the order of the attributes in terms of satisfaction/importance levels. In each model, the parameter associated with the attribute that has the lowest level of satisfaction/importance is set at 0, allowing all other parameters to be positive, assisting parameter interpretation. Specifically, the value of a parameter identifies its position in the satisfaction/importance scale with the larger parameter being representative of a more satisfied/important attribute.

The remaining column show the values of the parameters of the OL model. The OL model is completed with the constant term and the threshold parameters, as was explained in section 4. The dependent variable for this model is the overall satisfaction (OS) and the independent ones are all the attributes evaluated using the conventional rating scale. The values of the parameters in the OL model show how much each attribute contribute to explain the customer overall satisfaction with the service (i.e. OS).

For brevity, estimation results from the standard logit and order logit models are analysed in this article. These models assumes homogeneity in preference among individuals by estimating average parameters for the sample. This homogenous assumption is relaxed by estimating random parameter (RP) models (estimation results shown in the Appendix A). For brevity purpose, the comparison of alternative modelling approaches is conducted using standard models since the main findings, discussed below, still stand in light of the RP modelling results.

Tabla 43 Artículo 4: Table 3 – Estimation results of BW and ordered logit models (t-value in parentheses)

Variable	MNL- BW (Satisfaction)	MNL - BW (Importance)	Ordered Logit
Use of hybrid buses	1.326 (10.96)	1.083 (8.66)	0.039 (1.87)
Access time to bus stop	1.105 (9.34)	1.46 (11.88)	0.151 (2.80)
Egress time	1.054 (8.9)	1.630 (13.1)	0.067 (3.07)
Vehicle cleanliness	0.794 (6.84)	0.890 (7.36)	0.108 (2.29)
Ease of transfer	0.914 (7.84)	1.161 (9.41)	0.134 (4.24)
Information at stops	0.860 (7.27)	1.319 (10.65)	0.155 (2.86)
Information on board	0.455 (3.91)	0.249 (2.01)	0.059 (2.68)
Comfort of the buses	0.666 (5.71)	1.112 (9.02)	0.063 (2.63)
Service reliability / punctuality	0.763 (6.62)	2.336 (18.6)	0.270 (4.56)
Driver's kindness	0.480 (4.18)	0.624 (5.07)	0.021 (1.98)

Quality of bus stops	0.284 (2.41)	0.584 (4.67)	0.154 (2.76)
Information on mobile app	0.585 (4.93)	0.958 (7.68)	0.045 (2.23)
Line coverage	0.548 (4.67)	2.235 (17.56)	0.200 (3.59)
Information on the web page	0.494 (4.4)	0 (0)	0.101 (3.35)
Priority seats for PRM	0.406 (3.45)	1.621 (12.89)	0.055 (2.47)
Waiting time	0.375 (3.21)	1.988 (15.76)	0.142 (3.14)
Crowding level	0.209 (1.81)	1.478 (12.03)	0.145 (3.28)
Readability of map design	0.114 (0.98)	1.118 (8.97)	0.180 (3.25)
In-vehicle travel time	0.434 (3.66)	2.239 (17.29)	0.080 (3.71)
Service frequency and timetables	0.305 (2.62)	2.260 (17.82)	0.298 (5.11)
Driving style	0.075 (0.65)	1.474 (11.8)	0.078 (2.41)
Price / Fare	0 (0)	1.860 (15.03)	0 (0)
Calefaction / air conditioning	0.050 (0.43)	0.430 (3.49)	0.062 (2.86)
Noise	0.069 (0.6)	0.472 (3.87)	0.030 (2.66)
Constant	-	-	-1.305 (-2.88)
Mu(01)	-	-	1.427 (11.07)
Mu(02)	-	-	3.628 (33.61)
Mu(03)	-	-	6.997 (46.47)
Log-likelihood	-5817.674	-5373.256	-818.697
AIC/N	4.819	4.452	2.072
McFadden PseudoR2	-	-	0.11

As stated in the BW satisfaction model, users are very satisfied with the company's environmental policy when deploying hybrid buses (HY). In addition, users are also very satisfied with the access times to the stops (AT) and the egress times (DT), this mean that bus users in Santander see a good spatial coverage. As for the less satisfactory attributes, the price of tickets (PR) can be found as the least satisfactory. The current tickets fares vary depending on user type and the payment system used; however, the price is not higher than other public transport services nearby. The result may suggest the existence of a strategic voting behaviour, in the sense that the respondents strategically voted down their satisfaction with transport fares to reduce the chance that operators may increase fares in the future. This behaviour was also observed in previous studies conducted in the same city (Echaniz et al., 2018). The environmental characteristics such as noise (NO) and heating systems (CA) also show low levels of satisfaction, as well as several attributes

related to comfort during the trip: driving style (DS) and crowdedness (OC). Analysing the existing information channels in the service, it can be observed that the users are very satisfied with the information offered at stops, and somewhat less with the information available in mobile applications. While it is shown that for the remaining information sources (information on board the buses and on the website) users are not satisfied.

The importance BW based model shows that the most important attributes are those directly related to the basic characteristics of the service such as service reliability (SR), frequency (SE), on-board travel time (TT) and coverage of the lines (LC). Conversely, the least important attributes are noise level (NO), air conditioning systems (CA), information on board (IB) and on the website (IW).

According to the OL model, the attributes that show the high parameter values are service frequency (SE), service reliability (SR) and coverage of the lines (LC). By contrast, the driver kindness (DK), noise level (NO) and the use of hybrid technologies (HY) are the ones with the lowest parameter estimates. The threshold parameters show a nonlinearity in different rating points, which means that from the user's perspective, difference levels of effort are required to improve the service by one satisfaction point, such as from very bad to bad vs. from good to very good. These results are consistent with the findings obtained in the previous study developed in the same city (Echaniz et al., 2018).

With the data available we wanted to analyse if there is any connection between the models derived from the Best-Worst exercise and the results obtained from the Ordered Logit modelling and satisfaction ratings. Tabla 44 shows the correlation in parameters obtained from the two approaches. The correlation between the averages of the traditional satisfaction rating for each attribute and the BW satisfaction model is nearly perfect, with correlation coefficient of 0.95. The ordered Logit model shows a considerable correlation ($r = 0.486$) with the BW model of importance.

Tabla 44 Artículo 4: Table 4 – Correlation coefficients

	Importance_BW	Satisfaction_BW	Ordered Logit	Satisfaction Rating
Importance_BW	1			
Satisfaction_BW	0.083	1		
Ordered Logit	0.486	0.074	1	
Satisfaction Rating	-0.056	0.946	0.016	1

A deeper investigation into these strong correlations is presented in Figura 15 and Figura 16. Parameter values differ in scale from one model to another, and thus a direct comparison of parameter estimates does not show the true correlation between them. Therefore, both sets of the parameter estimates were standardized, providing some positive and some negative standardized scores showed in Figura 15 and Figura 16. The first comparison has been made between the two values representing the satisfaction. On the one hand, we have the information based on the classic revealed preference survey rating, obtaining the average satisfaction with values between 0 and 4. On the other hand, the MNL model based on the BW satisfaction data. As can be seen in Figura 15, the

correlation between these two values is considerably high. Most of the attributes show a similar tendency for both values although there are a few exceptions such as information in mobile apps (IM), line coverage (LC), information on board (IB) and quality of the stops (ST). Therefore, both methods lead to the same results, lending support to the hypothesis that BW method can replace the traditional satisfaction rating.

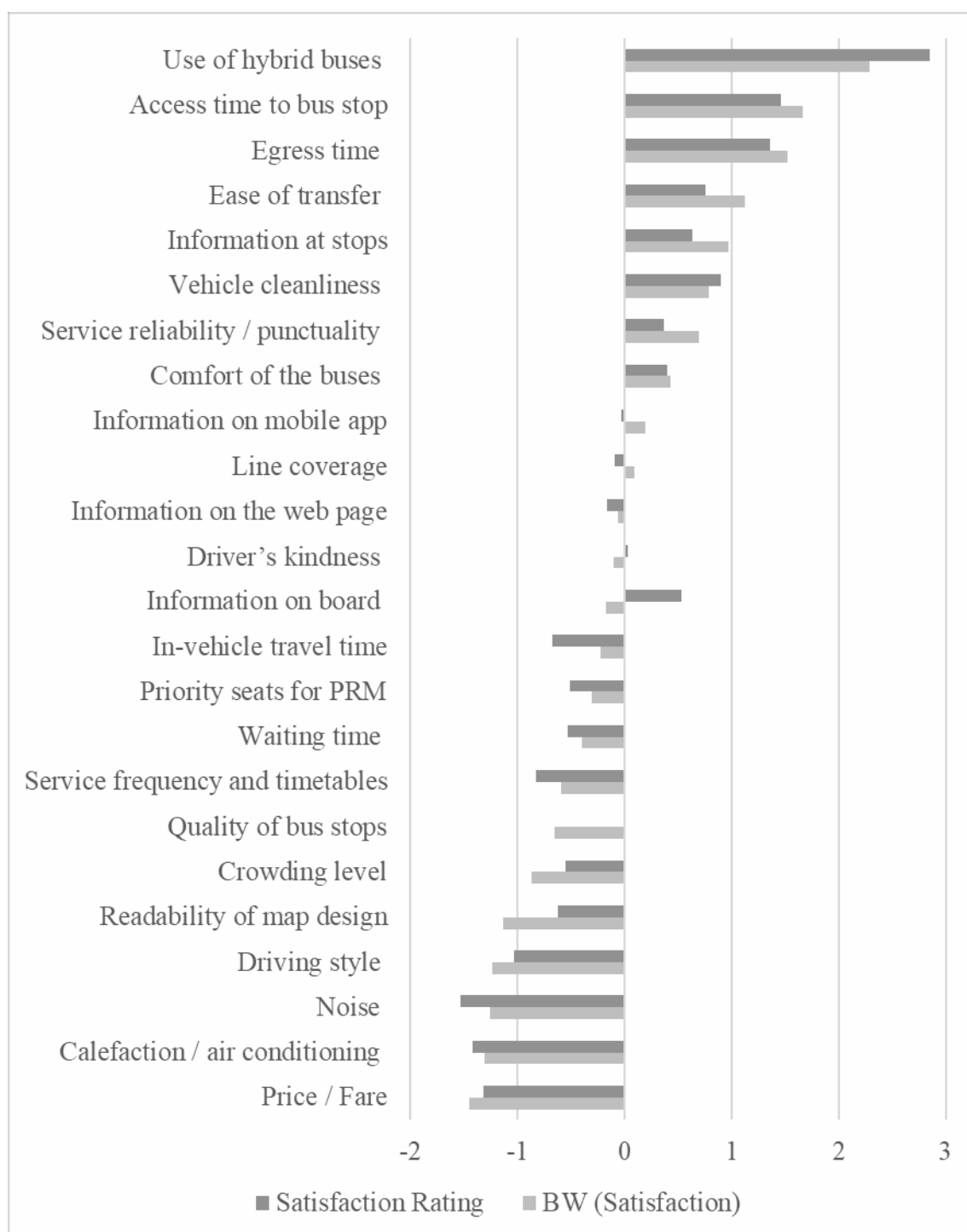


Figura 15 Artículo 4: Figure 2 - Comparison between BW (satisfaction) model and satisfaction rating

For the comparison of the importance, the Ordered logit and MNL models based on the importance BW have been selected. The value that a parameter has in an Ordered model can be associated with the weight it has when explaining the dependent variable. In other words, the parameter explains the contribution of each attribute to the overall satisfaction. An analysis of these two set of parameters, presented in Figura 16, shows a lower level of correlation than previous satisfaction one. Even the level of importance of the most important attributes shows a similar trend in both models, the rest of the attributes show very small correlation. In consequence, although both models represent a certain level of importance of the variables, these values are not the same and, therefore, represent different importance concepts.

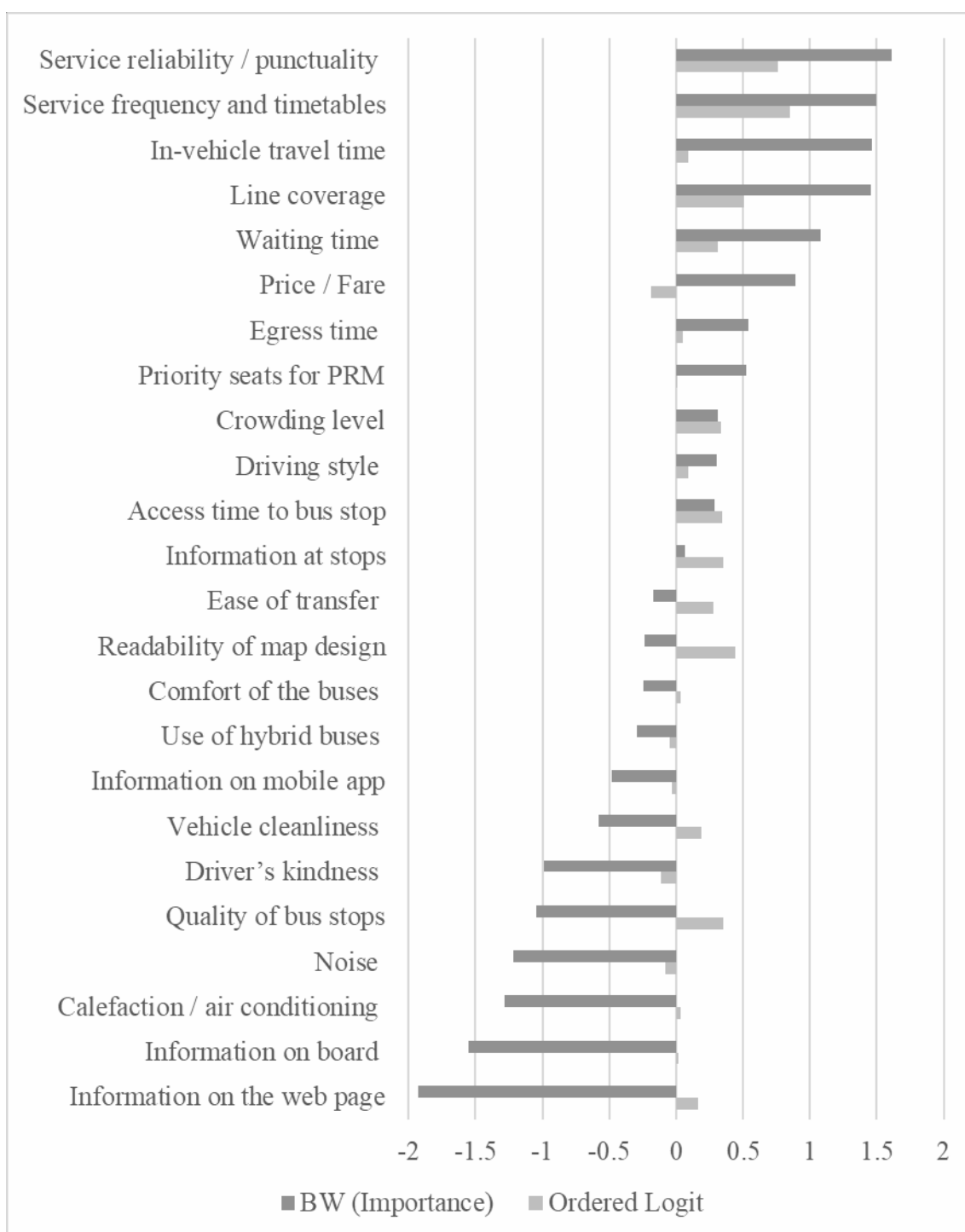


Figura 16 Artículo 4: Figure 3 – Comparison between BW (importance) and Ordered Logit models

5.2. Importance-Performance analysis

The importance-performance analysis (IPA) (Martilla and James, 1977) is a widely used decision tool. The basis of this method is to cross both result (performance level and importance) in the same graph. Four quadrants are defined, each of them with a different level of improvement priority. The four quadrants are typically identified as ‘keep up the good work’ (Q1 – “Important and satisfied”), ‘possible overkill’ (Q2 – “Not important and satisfied”), ‘low priority’ (Q3 – “Not important and not satisfied”) and ‘concentrate here’ (Q4 – “Important and not satisfied”) (Sever, 2015). The attributes on the Q1 are

considered the strengths of the service, attributes that are performing well and where investments should be kept equal to maintain the satisfaction level. Attributes on the Q2 contain attributes that are not important for users but still are performing strongly, that means that there is a possible waste of resources used in these attributes. Attributes on the Q3 are the ones with the lowest level of priority for investment. Finally, Q4 shown the main improvement priorities of the service, attributes that are important for users but are not performing good enough to satisfy the customers.

IPA method was recently used in the literature to compare explicit and implicit importance on transit services (Cao and Cao, 2017). Explicit importance is obtained by asking the respondent, either through a conventional rating method or some other methods, while implicit importance is derived from an OL model. They found that the priorities for service improvement based on explicit importance are different from those based on implicit importance. Aiming to verify this important finding, Figura 17 presents the Importance – Performance Analysis (IPA) in which explicit importance and satisfaction obtained from BW model are compared with those obtained from the conventional method (rating and OL). Satisfaction levels are represented in the horizontal axis, while importance levels are shown on the vertical axis. Blue dots show the positions of the service attributes according to the BW data and modelling results. Orange dots show the same but using the conventional rating as satisfaction and OL parameters as importance. The values are normalized for presentation purposes.

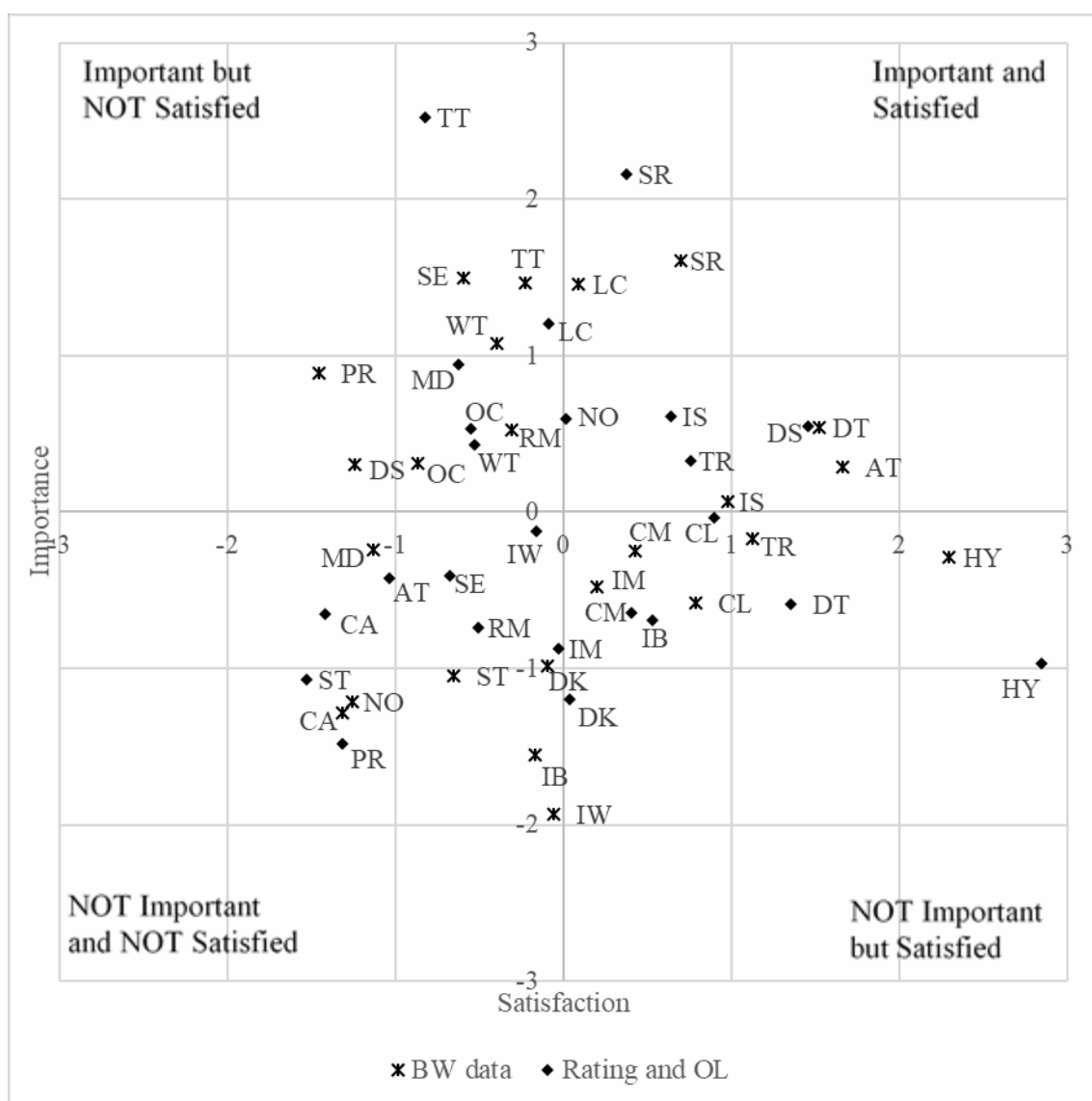


Figura 17 Artículo 4: Figure 4 – Importance – performance analysis

Close to 55% of the service attributes position themselves in a different quadrant when using a conventional satisfaction rating vs. best-worst rating. For some attributes, such as line coverage (LC) and information on mobile phones (IM), the quadrants are different but positions are very close to each other. For some other attributes, such as ticket fare (PR), the differences are quite remarkable in which the BW method identifies ticket fare as an important variable while the conventional OL model suggests the opposite (PR is the not important).

The differences are much greater in importance than in satisfaction. Moreover, satisfaction values are quite similar between the two methods, according to the positions relative to the horizontal axis. Most of the attributes are placed in a very close horizontal axis value, which means a similar satisfaction level. Quadrant position differences are therefore a result of different important levels identified by the alternative methods. Given that attribute importance derived from BW model is explicit (since users explicitly choose the most and least important attributes from a set of attribute) and that of the OL model is implicit (deriving from parameter estimates of an overall satisfaction model), we

support Beck and Rose (2016) argument that BW scaling is better than conventional rating for defining attribute importance.

Using results of the BW methods, a change to the current fare structure may result in a higher level of satisfaction since PR is highly important but is the most unsatisfied attribute. However, Results of the ordered model, suggests that any change in the fare policy would not generate increase the overall satisfaction of the service. Such the conflicting evidence is applied also to the service frequency in which the BW method suggests improving service frequency would improve customer satisfaction while the OL method identifies service frequency as a low priority. These differences suggest that the concept of importance may differs between the two modelling methods. We further investigate this in the next section.

5.3. Deriving attribute-specific satisfaction from BW models

An outstanding question regarding the use of BW survey method in customer satisfaction study is that whether it would be possible to obtain customer satisfaction level with each service attribute from the BW survey methods. This section addresses this question by performing a regression analysis to estimate the average level of satisfaction with each attribute based on the overall level of satisfaction and the BW model parameters. Two models are estimated: one for satisfaction level and one for the importance level of each attribute.

The BW satisfaction model shown in section 5.1 represents the relative satisfaction level of the many attributes that together define the entire service. Thus, these parameters are best interpreted as how much more or less an average customer is satisfied with a certain service attribute, compared to the reference attribute. That challenge is to convert this relativity of satisfaction level to an average level of satisfaction for each attribute, which is usually available in the traditional rating survey with specific questions. We propose a way to supplement this information for BW method by adding a constant term to the regression model, effectively converting the BW model parameters to the satisfaction level for each of the attributes included in the BW survey. More specifically, it is the Overall Satisfaction that acts as the constant term, as it remains constant for the whole sample. The regression model is specified in equation (60) where the dependent variable is the satisfaction (OS) rating for each attribute and the independent variables are the Overall Satisfaction (OS) and the BW satisfaction model parameters.

$$SatisfactionRating_i = OverallSatisfaction + e^{\delta_k^{BW(Satisfaction)}} \quad (60)$$

Tabla 45 Artículo 4: Table 5 – Satisfaction regression model

	Parameters	t-value
Overall Satisfaction	0.774	52.017
exp[BW(Satisfaction)]	0.301	14.397
R ²	0.999	
Adjusted R ²	0.954	

Tabla 45 shows the estimation results, confirming that it is possible to accurately estimate the average satisfaction level for each of the attributes from the BW data. The model has an R^2 of 0.999, indicating that nearly 100% of the variation in attribute-specific satisfaction is explained by the OS level and BW model parameters (i.e., the relativity satisfaction of different attributes). The results suggest that conventional rating surveys can be replaced by BW scaling surveys for measuring attribute satisfaction.

The parameters associated to an Ordered model define up to a certain level the implicit importance of the different attributes. As shown in section 5.1, the implicit importance of the OL model and the explicit importance derived from the BW MNL model are correlated, although the correlation is not as strong as in the case of satisfaction. Tabla 46 shows the result of the regression model defined in equation (61).

$$\beta^{OL} = SatisfactionRatings_i + e_k^{\delta^{BW(Importance)}} \quad (61)$$

Tabla 46 Artículo 4: Table 6 – Importance regression model

Parameters	t-value
Satisfaction ratings	0.017 1.914
Exp[BW(Importance)]	0.015 3.282
R^2	0.788
Adjusted R^2	0.733

The goodness of fit indicators shows that the parameters of the OL model can be estimated using BW data to a certain extent. This means that there is a difference in how importance is captured in each method. In BW method, importance is considered explicit while in OL it is implicit, and thus, they are related but not the same. The parameters of an OL model encapsulate both the explicit importance and the satisfaction. In other words, the parameters of the OL model represents not only the importance that each attribute as a separate factor has within the system, but also the satisfaction with the entire system. This explains why there is such a difference in the IPA analysis shown in section 5.2, as the OL goes further than simply defining a level of importance of the different attributes.

6. CONCLUSIONS

This study showed that the conventional method (rating and ordered logit models) and the BW methods are both suitable for analysing users' satisfaction with public transport services. Correlation and modelling analysis results indicates that conventional rating and BW methods are equivalent when studying the satisfaction levels of the different attributes of the service. The regression model shows that it is possible to reproduce the rating results by using BW data, and thus, BW method can replace the conventional rating method. This is an important finding which has a significant implication for improving the efficiency of customer satisfaction surveys and bringing positive effect to the respondents. With the BW survey methods being less time consuming and easier for the respondent to answer than the traditional rating method, this finding suggests that we can replace the lengthy and repetitive customer survey with a series of games presented as

best-worst choice tasks, especially when the results reported in this paper could be replicated on different datasets and/or in different settings.

The roles that different service attributes play in explaining customer overall satisfaction turn out to be very different, depending on the modelling methods. This finding is consistent with the growing evidence on the difference between explicit importance and implicit importance in customer satisfaction studies. Specifically, important drivers of customer satisfactions obtained from the best-worst explicit importance levels are travel time, service frequency and price while these key service attributes did not come out as important factors based on the OL implicit importance levels. In this sense, the BW scaling appears to be a good indicator of attribute importance, while the OL model parameters goes further than just defining an importance level. The regression model has shown that OL model parameters are not only influenced by the importance of the attributes but also by their satisfaction level.

The Importance-Performance Analysis (IPA) offers a new way to classify service attributes according to their importance/satisfaction levels. This helps transport operators and authorities to identify the key service attributes to improve. Again, the conclusions can be considerably different and sometimes opposite, depending on the adopted method. The main differences between the two methods are ticket fare (PR) and service frequency (SE). The main differences are placed in importance levels, as satisfaction results are very similar between both methods. The importance levels derived from BW data are in line with the literature. In conclusion, the explicit importance levels obtained by using the BW method are more accurate than the implicit importance derived from the OL model. Therefore, the IPA based on BW scaling is a better indicator of which attributes should be the real priorities for operators. This finding verify results of Cao and Cao (2017), who concluded that improvement priorities based on implicit importance (OL model) were more reliable than those based on the explicit conventional rating.

Although the results of this study show that BW scaling can effectively replace the conventional rating method, several considerations should be taken into account when implementing the BW scaling method for customer satisfaction surveys. First, to obtain the average satisfaction rating of the attributes by using BW data, it is necessary to fit the regression model that connects both models, therefore, a preliminary study is required to estimate the regression model, similar to the study carried out in this article. In addition, the Overall Satisfaction of the service must be rated independently to the method used.

Regarding to future research, results reported in this paper suggest that correlations between the means of random parameter estimates are as strong as correlations between non-random model parameters; however, there are differences in the deviations of the random parameters in the sense that less preference heterogeneities were found in the BW data than in the rating data. This might suggest that BW survey method results in less noise in the data than the traditional rating survey method does. More research is required to verify this initial finding, such as using different datasets and/or parsimonious models where preference heterogeneities are segmented into systematic vs random heterogeneity. Also, different modelling approaches should be tested to fit the BW data, for example the repeated best-worst model. In addition, in an ongoing research we are investigating the extent to which customer satisfaction varies across the service levels during the whole

day. Automatic Vehicle Location data is being used to identify not only the line but also the bus the respondents were on.

ACKNOWLEDGMENTS

We greatly appreciated the three anonymous referees' constructive comments which have materially improved the paper. This study has been possible thanks to the financing of the Spanish Ministry of Economy and Industry in the TRA2015-69903-R Project (co-funded with ERDF funds), to the Spanish Ministry of Science, Innovation and Universities through the project TRA2017-85853-C2-1-R, to the training grant FPU15 / 02990 of the Spanish Ministry of Education, Culture and Sports, to European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement No 688082 - SETA Project) and to the Full Professor Angel Ibeas, director of the Transport Systems Research Group.

APPENDIX A

A.1 Random parameter models

Tabla 47 Artículo 4: Table 21- Random parameter Ordered Logit

Attribute	Parameter	t-value
Non-Random Parameters		
Constant	-3,616	-5,35
Egress time	0,085	3,65
Vehicle cleanliness	0,221	3,45
Driver's kindness	0,057	3,75
Readability of map design	0,338	4,62
Noise	0,098	6,02
Service reliability / punctuality	0,576	7,17
Information on mobile app	0,071	2,93
Line coverage	0,412	5,51
Means for random parameters		
Use of hybrid buses	0,089	3,27
Access time to bus stop	0,340	5,67
Ease of transfer	0,311	6,79
Information at stops	0,333	4,76
Information on board	0,184	5,80
Comfort of the buses	0,137	3,98
Quality of bus stops	0,333	4,58
Information on the web page	0,213	5,14
Priority seats for PRM	0,139	4,45
Waiting time	0,301	4,97
Crowding level	0,332	5,18
In-vehicle travel time	0,211	6,59
Service frequency and timetables	0,566	7,57
Driving style	0,218	4,69

Calefaction / air conditioning	0,131	4,72
Scale parameters for dist. of random parameters		
Use of hybrid buses	0,063	9,43
Access time to bus stop	0,281	9,73
Ease of transfer	0,261	9,79
Information at stops	0,498	12,64
Information on board	0,201	10,74
Comfort of the buses	0,218	11,30
Quality of bus stops	0,150	6,31
Information on the web page	0,257	11,97
Priority seats for PRM	0,086	9,64
Waiting time	0,230	9,26
Crowding level	0,222	7,87
In-vehicle travel time	0,152	8,85
Service frequency and timetables	0,262	7,11
Driving style	0,152	13,54
CA	0,123	7,43
Threshold parameters		
Mu(01)	2,507	6,51
Mu(02)	6,835	14,27
Mu(03)	14,143	20,07
Log-likelihood		
AIC/N		
McFadden Pseudo R2		

Tabla 48 Artículo 4: Table 22 Mixed Logit model for Satisfaction BW data

Attribute	Parameter	t-value
Means for random parameters		
Access time to bus stop	1,170	4,73
Waiting time	0,714	2,92
In-vehicle travel time	1,218	0,79
Egress time	1,054	8,18

Ease of transfer	0,910	6,94
Service frequency and timetables	0,307	2,2
Service reliability / punctuality	0,768	5,84
Line coverage	0,545	4,02
Information at stops	0,862	6,56
Information on the web page	0,493	3,94
Information on board	0,455	3,44
Crowding level	0,208	1,52
Calefaction / air conditioning	0,052	0,38
Priority seats for PRM	0,418	3,06
Comfort of the buses	0,670	5,12
Vehicle cleanliness	0,795	6,09
Driving style	0,072	0,53
Driver's kindness	0,478	3,61
Use of hybrid buses	1,328	10,06
Noise	0,071	0,53
Information on mobile app	0,587	4,24
Quality of bus stops	0,285	2,06
Readability of map design	0,114	0,82
Scale parameters for dist. Of random parameters		
Access time to bus stop	0,050	0,3
Waiting time	0,305	1,67
In-vehicle travel time	1,496	0,51
Egress time	0,131	1,43
Ease of transfer	0,074	0,82
Service frequency and timetables	0,050	0,5
Service reliability / punctuality	0,017	0,21
Line coverage	0,051	0,55
Information at stops	0,005	0,05
Information on the web page	0,024	0,29
Information on board	0,098	1,17
Crowding level	0,059	0,67
Calefaction / air conditioning	0,034	0,35
Priority seats for PRM	0,122	1,24
Comfort of the buses	0,100	1,06
Vehicle cleanliness	0,019	0,22
Driving style	0,058	0,54
Driver's kindness	0,017	0,2
Use of hybrid buses	0,066	0,73
Noise	0,047	0,49
Information on mobile app	0,091	0,91
Quality of bus stops	0,016	0,17
Readability of map design	0,031	0,3
Log-likelihood	-19.636	

AIC/N	16,239
-------	--------

Tabla 49 Artículo 4: Table 23 Mixed Logit model for Importance BW data

Attribute	Parameter	t-value
Non-random parameters		
Egress time	1,628	11,47
Service reliability / punctuality	2,337	17,43
Line coverage	2,235	16,57
Vehicle cleanliness	0,889	6,35
Driver's kindness	0,625	3,93
Noise	0,470	3,06
Information on mobile app	0,956	6,56
Readability of map design	1,113	7,59
Means for random parameters		
Access time to bus stop	1,076	4,07
Waiting time	1,808	8,72
In-vehicle travel time	2,810	2,29
Price / Fare	1,858	13,59
Ease of transfer	1,149	7,97
Service frequency and timetables	2,260	16,91
Information at stops	1,315	9,25
Information on board	0,245	1,48
Crowding level	1,489	10,78
Calefaction / air conditioning	0,425	2,76
Priority seats for PRM	1,619	11,47
Comfort of the buses	1,124	7,79
Driving style	1,471	10,19
Use of hybrid buses	1,081	7,76
Quality of bus stops	0,582	3,94
Scale parameters for dist. Of random parameters		
Access time to bus stop	0,302	1,7
Waiting time	0,163	1,25
In-vehicle travel time	1,092	0,47
Price / Fare	0,029	0,36
Ease of transfer	0,061	0,65
Service frequency and timetables	0,062	0,78
Information at stops	0,027	0,32
Information on board	0,147	1,14
Crowding level	0,142	1,9
Calefaction / air conditioning	0,024	0,22
Priority seats for PRM	0,014	0,17
Comfort of the buses	0,138	1,46
Driving style	0,035	0,37
Use of hybrid buses	0,044	0,52
Quality of bus stops	0,019	0,18

Log-likelihood	-19.636
AIC/N	16,239

A.2 Comparison of RP models

Tabla 50 Artículo 4: Table 24 correlation coefficients of means of RP models

	Importance_BW	Satisfaction_BW	Ordered Logit	Satisfaction Rating
Importance_BW	1			
Satisfaction_BW	0,248	1		
Ordered Logit	0,417	0,060	1	
Satisfaction Rating	-0,102	0,806	0,007	1

Tabla 51 Artículo 4: Table 25 Correlation coefficients of scale parameters of RP models

	Importance_BW	Satisfaction_BW	Ordered Logit	Satisfaction Rating
Importance_BW	1			
Satisfaction_BW	0,948	1		
Ordered Logit	0,203	0,053	1	
Satisfaction Rating	-0,105	-0,062	0,075	1

“Todos los modelos son aproximaciones. Las suposiciones, ya sean implícitas o claras, nunca son exactamente ciertas. Todos los modelos están equivocados, pero algunos modelos son útiles”

“All models are approximations. Assumptions, whether implied or clearly stated, are never exactly true. All models are wrong, but some models are useful.”

(Box et al., 2009)

CAPÍTULO 5: RESULTADOS Y DISCUSIÓN

5.1 ENCUESTA DE SATISFACCIÓN TUS 2015

5.1.1 Resultados socioeconómicos de la muestra

Las encuestas de satisfacción se realizaron a lo largo de 15 días laborables en los meses de abril y mayo de 2015. La forma de encuesta utilizada fue la de entrevistas cara a cara a bordo de los autobuses. En el caso de no poder completar la encuesta en el tiempo que el usuario se encontraba a bordo del autobús se establecían 2 opciones: desechar la encuesta por no estar finalizada o apearse con el usuario para finalizar la encuesta en la parada y posteriormente esperar al siguiente autobús. En ambos casos, la eficiencia del proceso de encuestado se veía reducida considerablemente. La muestra mínima necesaria se estableció en 700 encuestas completadas, siguiendo el método de cálculo mostrado en el apartado 3.4 de esta tesis. Al final, se consiguieron realizar un total de 747 encuestas efectivas alcanzando un ratio de respuestas de unas 4 encuestas completadas por hora por encuestador.

La muestra final se compuso de un 71% de mujeres, las cuales se encontraban ligeramente sobre representadas. Dos terceras partes de los encuestados fueron menores de 44 años y cerca de la mitad (49%) de ellos trabajadores. Los estudiantes representaron un cuarto de la muestra. Un 60% contestaron afirmativamente a la posesión del carnet de conducir, sin embargo, solo el 40% disponía de vehículo propio. En cuanto al uso del transporte público, la mitad de los encuestados eran usuarios habituales, considerando usuarios habituales aquellos que utilizan el servicio entre 5 y 15 veces por semana. El motivo principal para el uso del transporte público fue el comúnmente entendido como obligatorio, esto es, trabajo o estudios, y en la mayoría de casos relacionados con el hogar. La gran mayoría de los encuestados (95%) utilizaban ya la tarjeta de transporte sin contacto para pagar el billete del autobús. Por último, el usuario mayoritario en el transporte de Santander se podía asociar a una persona con un salario medio-bajo. Sin embargo, este dato no es concluyente puesto que, al tratarse de una respuesta opcional, debido a que la gente suele ser reacia a contestarlo, casi un 40% de la muestra decidió no responderla. La descripción completa de la muestra puede observarse en la Tabla 52.

Tabla 52 Resultados socioeconómicos de la encuesta de satisfacción de 2015

Género	Hombre	29%
	Mujer	71%
Edad	<25	28%
	25-34	16%
	35-44	22%
	45-54	16%
	55-64	11%
	>65	7%

Situación laboral	Trabajador	49%
	Desempleado	17%
	Estudiante	25%
	Jubilado	9%
Carnet de conducir	Si	59%
	No	41%
Vehículo propio	Si	40%
	No	60%
Frecuencia de uso del autobús	< 5 viajes/semana	29%
	5 - 15 viajes/semana	50%
	15 - 30 viajes/semana	18%
	> 30 viajes/semana	3%
Motivo del viaje	Casa	32%
	Trabajo	22%
	Estudios	13%
	Sanidad	4%
	Compras	7%
	Ocio	11%
	Otros	11%
Modo de pago	Tarjeta de transporte	95%
	Efectivo	5%
Salario mensual	Bajo (< 900€)	31%
	Medio (900 - 1500€)	23%
	Alto (1500 - 2500€)	7%
	Muy alto (> 2500€)	1%
	Desconocido (sin respuesta)	38%

5.1.2 Resultados de la valoración de los atributos

Con el objetivo de facilitar la interpretación, los resultados de las valoraciones cualitativas de la encuesta se muestran de forma numérica en la Tabla 53. Los valores están comprendidos entre el valor 0 (“Muy Mal”) y 4 (“Muy Bien”). La satisfacción general con el servicio de transporte público era considerada buena, consiguiendo una puntuación de 2.81 sobre 4.

Tabla 53 Resultados de la valoración de la encuesta de satisfacción de 2015

Grupos	Atributos	Acrónimo	promedio	Desviación estándar	Moda
Nivel de Servicio	Tiempo caminando hasta la parada.	AT	2.91	0.80	3
	Tiempo de espera en la parada.	WT	2.45	0.89	3
	Tiempo de viaje.	TT	2.62	0.74	3
	Tiempo desde la parada final hasta el destino real.	TD	2.87	0.70	3
	Precio de los billetes.	PR	2.07	0.98	2
Servicio Ofertado	Facilidad de transbordos.	TR	2.67	0.80	3
	Servicios ofertados (Horarios, frecuencias).	SE	2.49	0.82	3
	Fiabilidad del servicio.	SR	2.76	0.73	3
	Líneas especiales para: fútbol, semana grande,...	EL	2.35	0.73	2
	Servicio nocturno / durante el fin de semana.	NS	2.16	0.83	2
	Cobertura de las líneas (itinerarios, paradas,...).	LC	2.63	0.79	3
	Información en las paradas y paneles informativos.	IS	2.77	0.83	3
Información	Información en soporte informático.	IWM	2.88	0.81	3
	Información en los monitores del autobús.	IB	2.59	0.74	3
	Ocupación.	OC	2.39	0.81	3
Confort	Calefacción / Aire acondicionado.	CA	2.44	0.78	3
	Espacio para personas de movilidad reducida.	RM	2.57	0.86	3
	Confort y comodidad de la flota de autobuses.	CM	2.73	0.59	3
	Limpieza de los autobuses.	CL	2.86	0.66	3
	Posibilidad de portar objetos/bultos, animales,...	OB	2.38	0.79	3
	Forma de conducción (Arranque, frenado,...).	DS	2.67	0.79	3
	Amabilidad del conductor.	DK	2.65	0.81	3
Sostenibilidad	Implantación de autobuses híbridos.	HY	3.06	0.75	3
	Contaminación acústica.	NO	2.27	0.65	2
	Satisfacción General	OS	2.81	0.63	3

Los resultados muestran que los usuarios estaban generalmente satisfechos con el servicio y con todos los aspectos que lo describen. El atributo que se considera peor es la tarifa. Este hecho puede entenderse como una respuesta estratégica, ya que los usuarios no suelen evaluar bien este atributo por temor a un posible aumento de las tarifas del servicio, sin embargo, el valor medio muestra que no se considera un factor insatisfactorio para los usuarios. Por el contrario, el atributo mejor valorado es el uso de autobuses híbridos. Cualquier acción asociada a una mejora ambiental del servicio es generalmente acogida con buenos ojos por los usuarios. Un aspecto que es valorado positivamente por los usuarios es la densidad de paradas que ofrece el servicio, puesto que tanto el Tiempo de acceso a la parada (AT) y el Tiempo desde la parada al destino final (ET) muestran un nivel de satisfacción alto. La desviación de las valoraciones es considerable, lo que hace indicar que existe heterogeneidad de opiniones en la muestra encuestada.

5.1.3 Modelización considerando variaciones sistemáticas y aleatorias en los gustos

Los modelos que se presentan a continuación permiten analizar aquellos atributos que mejor explican la variación de la satisfacción general del servicio. En esta primera encuesta se estimaron un total de 6 modelos, cada uno de ellos siendo una evolución del anterior. La modelización comienza con la estimación de un modelo base sencillo, que relaciona de forma lineal mediante un modelo ordered probit (véase apartado 3.2) la valoración de los atributos y la valoración general. Para en los sucesivos modelos añadir la variabilidad sistemática de los usuarios, la importancia derivada del ranking realizado en la encuesta y la variación aleatoria de la muestra. La notación exacta de los modelos se describe con mayor detalle en el apartado 3.5.1. Para todos ellos se utilizó la misma muestra compuesta por las 747 observaciones obtenidas en la campaña de encuestas.

Una gran parte de los atributos evaluados en la encuesta no resultaron ser estadísticamente significativos, por lo que no se han incluido en los modelos finales estimados, puesto que no mejoran la capacidad predictiva del modelo. Las variables relativas al confort de los autobuses y el Aire acondicionado / calefacción han sido combinadas en una variable única llamada Confort debido a su alto grado de correlación. Igualmente, por la misma razón, las variables relacionadas con la información (en plataformas digitales, a bordo del autobús y en la parada) han sido combinadas en la variable Información. Finalmente, debido a que ningún encuestado valoró la Satisfacción General del servicio con una puntuación de “Muy Mal”, las opciones de respuesta J del modelo se han reducido a 4.

5.1.3.1 *Modelo Base*

Para comprobar la consistencia del modelo es necesario verificar que los parámetros de las variables independientes del modelo muestren un nivel de significancia estadística suficiente y que los signos de los mismos sean los correctos. Un incremento en la calidad de uno de los atributos del servicio se asume que genera una mejora positiva en la Satisfacción General del servicio. Esto quiere decir que un incremento en un variable independiente (atributos) genera un incremento en la variable dependiente (Satisfacción general). Por lo tanto, los parámetros de todas las variables independientes deben ser positivos. Otra cuestión importante es que, debido a la base teórica del modelo (véase

apartado 3.2), el primer segmento de las opciones de respuesta se encuentra comprendido entre $-\infty$ y 0. Siendo todos los parámetros positivos no es posible obtener un valor negativo en el modelo considerando solo las variables independientes, por lo tanto, es necesario que la constante del modelo sea negativa para poder obtener un valor resultante del modelo negativo. En cuanto al nivel de significatividad estadística, el valor de test t muestra la significatividad de cada parámetro. Un valor absoluto de test t mayor a 1,96 muestra que un parámetro es estadísticamente significativo para un intervalo de confianza del 95%. La mayoría de los parámetros han resultado significativamente estadísticos salvo el precio de los billetes y el servicio ofertado. Ambos atributos se han mantenido en el modelo debido a su importancia en la gestión del servicio.

Debido a que todas variables independientes v_i y la variable dependiente q_i se encuentran expresadas en la misma unidad es posible realizar una comparativa directa entre los valores de los parámetros. Los valores de los parámetros representan el peso que cada atributo tiene a la hora de explicar la Satisfacción General del servicio, dicho de otra forma, el nivel de importancia de cada atributo. Mayor el parámetro, mayor la contribución de ese atributo a la Satisfacción general. La Tabla 54 muestra los resultados del modelo, se puede observar como el Confort es el parámetro con el valor más alto. En cuanto al resto de atributo, se puede observar que, con unos buenos tiempos de acceso a las paradas, una buena cobertura de las líneas con transbordos fáciles, con buena información y con una conducción suave se consigue una calidad del servicio muy alta.

El ajuste del modelo se determina utilizando el coeficiente Log-Verosimilitud, AIC/N y Count R^2 . Este primer modelo se ha utilizado como base para observar la mejora conseguida con la implementación de la heterogeneidad sistemática y aleatoria y la la importancia de los atributos contemplada en los sucesivos modelos. El valor más intuitivo es el valor de Count R^2 , que define el porcentaje de observaciones que el modelo es capaz de predecir correctamente. Esto es, este primer modelo que tiene un valor de 0,724 quiere decir que el modelo es capaz de predecir correctamente el 72,4% de las observaciones.

Tabla 54 Resultados de estimación del modelo base (BM)

BM		
F.Log-Verosimilitud	-497.064	
AIC/N	1.371	
Count R^2	0.724	
Variable	Coef.	Test t
Constante	-4.307	-11.53
Tiempo de acceso a la parada	0.248	4.01
Tiempo de espera	0.136	2.25
Tiempo de viaje	0.202	2.87
Precio del billete	0.078	1.54
Facilidad de transbordo	0.263	4.21
Servicio ofertado (Frecuencias)	0.064	0.98
Fiabilidad del servicio	0.178	2.52
Servicios nocturnos/fines de semana	0.132	2.22
Cobertura de las líneas	0.250	3.91
Información*	0.270	2.86

Confort **	0.579	6.22
Forma de conducción	0.285	4.38
Parámetros de límite		
μ_1	1.745	17.35
μ_2	4.606	33.67

* Promedio de las tres variables del grupo.

** Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.3.2 Modelo considerando la variación sistemática en los gustos de los usuarios (SVM)

El modelo estimado (Tabla 55) que considera la variación sistemática en los gustos de los usuarios muestra una mejora considerable en el ajuste del modelo. En este modelo se han mantenido las variables independientes presentes en el modelo base (BM) mostrado en el apartado 5.1.3.1, con una excepción, el precio de los billetes ahora solo se muestra interactuando con variables socioeconómicas. Dos características socioeconómicas han sido añadidas como variables independientes (motivo: trabajo y viajes > 30) por ser estadísticamente significativas. Los parámetros de ambas variables socioeconómicas muestran un valor negativo muy alto, lo que muestra que los usuarios que cogen el autobús para ir a trabajar y los usuarios más frecuentes del servicio tienen por defecto una valoración más negativa del mismo que el resto de usuarios.

La consistencia del modelo es adecuada de acuerdo las directrices definidas en el apartado 5.1.3.1. En este caso concreto, debido a la existencia de las interacciones, el criterio de no negatividad de los parámetros de las variables independientes debe cumplirse considerando la suma de todos los parámetros afectando a esa variable. Por ejemplo, en el caso de las interacciones Edad_25_35*Confort o Jubilado*Confort, el valor de las interacciones es negativa pero no suficientemente grande como para que la suma del parámetro independiente Confort y la interacción resulte negativa. Es necesario entender también, que ambas interacciones no pueden darse de forma simultánea, puesto que ningún usuario comprendido entre 25 y 35 años se encontraba jubilado. La edad oficial para jubilarse en la localización de la encuesta es de 65 años.

Como aporte interesante de este modelo, se puede observar que el motive del viaje representa un influyente importante a la hora de evaluar el Servicio. Puesto que todos los motivos disponibles en la encuesta se muestran en el modelo de forma independiente o interactuando con algún atributo.

Tabla 55 Resultados de estimación del modelo con interacciones (SVM)

SVM	
F.Log-Verosimilitud	-467.011
AIC/N	1.331
Count R ²	0.743

Variable	Coef.	Test t
Constante	-4.141	-10.38
Tiempo de acceso a la parada	0.203	3.05
Tiempo de espera	0.129	2.05
Tiempo de viaje	0.160	2.12
Facilidad de transbordo	0.211	3.05
Servicio ofertado (frecuencias)	0.060	0.88
Fiabilidad del Servicio	0.177	2.41
Cobertura de las líneas	0.162	2.21
Información *	0.324	3.27
Confort **	0.663	6.49
Forma de conducción	0.283	4.18
Motivo: Trabajo	-0.818	-2.18
Viajes>30	-1.972	-1.96
Interacciones		
Edad_25_35 * Tiempo de viaje	0.510	3.05
Edad_25_35 * Confort **	-0.514	-3.06
Trabajador * Servicios nocturnos / fin de semana	0.133	2.45
Estudiante*Facilidad de transbordo	0.129	2.25
Jubilado* Facilidad de transbordo	0.569	2.47
Jubilado* Confort **	-0.474	-1.98
Motivo: Casa * Precio del billete	0.177	2.99
Motivo: Trabajo * Cobertura de las líneas	0.427	3.08
Motivo: Estudios* Tiempo de espera	0.141	1.89
Motivo: Sanidad * Precio del billete	0.377	3.19
Motivo: Compras * Tiempo de acceso a la parada	0.131	1.89
Motivo: Ocio * Tiempo de acceso a la parada	0.147	2.32
Viajes_5_15* Facilidad de transbordo	0.071	1.95
Viajes>30* Forma de conducción	0.986	2.56
Ingresos< 900* Tiempo de acceso a la parada	0.080	2.21
Parámetros de límite		

μ_1	1.894	16.88
μ_2	4.921	31.74

*Promedio de las tres variables del grupo.

**Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.3.3 *Modelo con variaciones sistemáticas y variables ponderadas (SVMW)*

La ponderación de las variables significa la inclusión de la importancia explícita mostrada por cada usuario a los distintos atributos presentes en la encuesta. Al realizar la encuesta, los usuarios establecieron un ranking de importancia entre los distintos atributos pertenecientes aun un mismo grupo, ordenándolos de mayor a menor importancia. A su vez, al finalizar la encuesta los encuestados realizaron ese mismo ejercicio de ranking con los grupos que agrupaban los atributos. La ponderación se ha incluido en el modelo de acuerdo a lo definido en el apartado 4.1.2.

La consistencia del modelo estimado (Tabla 56) es correcta de acuerdo a las directrices establecidas en la sección 5.1.3.1. En este caso particular, la condición de no negatividad ha de cumplirse considerando las interacciones y las ponderaciones de forma conjunta.

Las variables ponderadas cambian su valor dependiendo de la importancia que un atributo tiene para cada usuario. Las variables simplemente ponderadas $\delta_{nk} \cdot v_{ik}$, la variable obtendrá el valor más alto cuanto el atributo sea considerado como el más importante dentro de su grupo $\delta_{nk} = 1$, donde v_{ik} representa la valoración de una variable k realizada por un individuo i . Por el contrario, si ese mismo atributo es considerado como el menos importante de su grupo, la variable asociada no se considerará en el modelo $\delta_{nk} = 0$. En el caso de la ponderación doble, $\delta_{nk} \cdot \delta_{mk} \cdot v_{ik}$, La importancia del atributo intercede con la importancia del grupo, de forma que el valor máximo se obtiene cuando el atributo es el más importante dentro de un grupo y a su vez, el grupo es el más importante de entre todos los grupos, $\delta_{nk} = 1$ y $\delta_{mk} = 1$. Sin embargo, en el caso de que el atributo o el grupo sean considerados los menos importantes en sus respectivos rankings, la variable será omitida del modelo, $\delta_{nk} = 0$ o $\delta_{mk} = 0$.

Cuando el parámetro obtenido para una ponderación es positivo, la calidad incrementa para un Usuario que considera esa variable importante, esto ocurre en el caso de $\delta_n \cdot \delta_m \cdot$ Precio del billete * Motivo: Sanidad. Sin embargo, cuando el parámetro es negativo, quiere decir que la contribución de esa variable a la Satisfacción General es menor a medida que la variable gana importancia.

Tabla 56 Resultados de estimación del modelo con interacciones y variables ponderadas (SVMW)

SVMW	
F.Log-Verosimilitud	-463.186
AIC/N	1.323
Count R ²	0.750

Variable	Coef.	Test t
Constante	-4.340	-10.84
Tiempo de acceso a la parada	0.288	4.2
Tiempo de espera	0.133	2.11
Tiempo de viaje	0.216	2.68
Facilidad de transbordo	0.204	2.94
Servicio ofertado (Frecuencias)	0.067	0.99
Fiabilidad del Servicio	0.204	2.78
Cobertura de las líneas	0.171	2.33
Información *	0.321	3.25
Confort **	0.70	6.8
Forma de conducción	0.319	4.73
Motivo: Trabajo	-0.857	-2.3
Interacciones		
Edad_25_35* Tiempo de viaje	0.474	2.83
Edad_25_35* Confort **	-0.472	-2.82
Trabajador * Servicios nocturnos / fin de semana	0.099	1.86
Estudiante * Facilidad de transbordo	0.114	1.98
Jubilado * Facilidad de transbordo	0.598	2.6
Jubilado * Confort **	-0.534	-2.23
Motivo: Trabajo * Cobertura de las líneas	0.454	3.27
Motivo: Estudios * Tiempo de espera	0.147	2.02
Motivo: Ocio * Tiempo de acceso a la parada	0.152	2.45
Viajes_5_15* Facilidad de transbordos	0.081	2.24
Variables ponderadas		
δ_n * tiempo de acceso a la parada	-0.131	-2.25
δ_n * δ_m * Tiempo de viaje	-0.160	-2.62
δ_n * Forma de conducción *Viajes>30	0.359	2.79
δ_n * δ_m * Precio del billete * Motivo: Casa	0.216	3.28
δ_n * δ_m * Precio del billete * Motivo: Sanidad	0.454	3.54
δ_n * δ_m * Tiempo de acceso a la parada * Motivo: Compras	0.428	2.93

Parámetros de límite		
μ_1	1.898	16.83
μ_2	4.946	31.52

*Promedio de las tres variables del grupo.

**Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.3.4 Modelo base con parámetros aleatorios (RBM)

La inclusión de parámetros aleatorios al modelo genera un modelo con tanto parámetros constantes como aleatorios. Para aceptar que un parámetro se distribuye aleatoriamente es necesario que tanto la media como el parámetro de escala sean estadísticamente significativos y que el signo de la media sea correcto. En este modelo (Tabla 57) los parámetros considerados como aleatorios han sido el Precio del billete, La Facilidad de transbordo, la Información y el Confort. La mejora ocasionada por la inclusión de los parámetros aleatorios se discute más en detalle en el apartado 5.1.4.

Las medias de los parámetros aleatorios muestran valores similares al modelo de parámetros constantes (apartado 5.1.3.1). El valor de la media puede variar de forma considerando el parámetro de escala de la distribución. En este caso, todos los parámetros se han considera distribuidos siguiendo una distribución normal, por lo que el parámetro de escala representa la desviación estándar de la distribución. El precio de los billetes muestra la mayor variación con respecto a la media, mostrando la heterogeneidad en la valoración de ese atributo. El confort, aun mostrando un parámetro de escala estadísticamente significativo, el valor del mismo es muy pequeño (0.0448) en comparación al valor de su media (0.754).

Tabla 57 Resultados de estimación del modelo base con parámetros aleatorios (RBM)

RBM		
F.Log-Verosimilitud	-494.197	
AIC/N	1.374	
Variable	Coef.	Test t
Constante	- 5.570	-12.40
Tiempo de acceso a la parada	.307	4.32
Tiempo de espera	.191	2.76
Tiempo de viaje	.246	3.30
Servicio ofertado (frecuencia)	.088	1.31
Fiabilidad del Servicio	.230	3.09
Servicio nocturno / fin de semana	.168	2.49

Cobertura de las líneas	.307	4.54
Forma de conducción	.361	5.24
Medias de los parámetros aleatorios		
Precio del billete	.108	1.98
Facilidad de transbordo	.338	4.92
Información *	.363	3.37
Confort **	.754	7.38
Parámetros de escala de las distribuciones de los parámetros aleatorios		
Precio del billete	.163	6.67
Facilidad de transbordo	.192	9.21
Información *	.183	8.99
Confort **	.045	2.25
Parámetros de límite		
μ_1	2.205	13.56
μ_2	5.956	22.63

*Promedio de las tres variables del grupo.

**Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.3.5 Modelo considerando la variación sistemática y aleatoria en los gustos de los usuarios (SRVM)

En el modelo que considera la variación sistemática y aleatorias en los gustos (Tabla 58), las diferencias con el modelo de solo variaciones sistemáticas (SVM) mostrado en el apartado **¡Error! No se encuentra el origen de la referencia.**, son que las interacciones a sociadas a parámetros aleatorios también resultan ser aleatorios. Al igual que en el molde RBM (apartado 5.1.3.4) todos los parámetros aleatorios se distribuyen siguiendo una función normal. La consistencia del modelo también es adecuada cumpliendo las condiciones de no negatividad de los parámetros y de niveles de significatividad estadística. Tanto como para los parámetros e interacciones constantes como para las aleatorias y sus respectivos parámetros de escala.

Los parámetros no aleatorios siguen la misma estructura que en modelos anteriores. Los parámetros aleatorios han sufrido una modificación por la inclusión de las interacciones, obteniendo de esta forma una heterogeneidad en la media de las distribuciones. Dicho de otra forma, las interacciones entre las variables socioeconómicas y las variables con parámetros aleatorios afecta al valor de las medias de los mismos. El Precio del billete muestra una media no significativa mientras que la desviación estándar y la heterogeneidad debido a las interacciones si lo son, por esta razón se ha mantenido en el

modelo como un parámetro aleatorio. Tanto en el caso del Precio del billete como en la Facilidad del transbordo, los valores de la heterogeneidad han resultado ser positivos. Mientras que, en el confort, en ambas interacciones con usuarios jóvenes y jubilados los valores de las medias se reducen considerablemente, reduciendo la importancia de ese atributo.

Tabla 58 Resultado del modelo que con variaciones sistemáticas y aleatorias en los gustos (SRVM)

SRVM			
F.Log-Verosimilitud		-464.444	
AIC/N		1.335	
Variable		Coef.	Test t
Constante		-5.734	-11.22
Tiempo de acceso a la parada		0.279	3.62
Tiempo de espera		0.219	2.89
Tiempo de viaje		0.208	2.47
Fiabilidad del Servicio		0.239	2.92
Cobertura de las líneas		0.232	2.81
Forma de conducción		0.393	5.12
Motivo: Trabajo		-1.182	-2.80
Viajes>30		-3.037	-1.97
Interacciones			
Edad_25_35* Tiempo de viaje		0.754	3.70
Trabajador* Servicio nocturno / fin de semana		0.198	3.14
Motivo: Trabajo * Cobertura de las líneas		0.588	3.66
Motivo: Estudios * Tiempo de espera		0.171	1.73
Motivo: Compras*Tiempo de acceso a la parada		0.183	2.26
Motivo: Ocio * Tiempo de acceso a la parada		0.194	2.50
Viajes>30* Forma de conducción		1.519	2.70
Ingresos< 900* Tiempo de acceso a la parada		0.112	2.56
Medias de los parámetros aleatorios			
Precio del billete		0.010	0.15
Facilidad de transbordo		0.287	3.47
Información *		0.454	3.85
Confort **		0.969	8.02

Parámetros de escala de las distribuciones de los parámetros aleatorios		
Precio del billete	0.187	7.09
Facilidad de transbordo	0.259	10.96
Información *	0.166	7.68
Confort **	0.112	5.09
Heterogeneidad en las medias de los parámetros aleatorios		
Motivo: Casa*Precio del billete	0.218	2.94
Motivo: Sanidad*Precio del billete	0.516	3.42
Estudiante * Facilidad de transbordo	0.199	2.81
Jubilado * Facilidad de transbordo	0.845	3.12
Viajes_5_15* Facilidad de transbordo	0.103	2.34
Jubilado * Confort **	-0.729	-2.57
Edad_25_35* Confort **	-0.768	-3.66
Parámetros de límite		
μ_1	2.592	12.88
μ_2	6.895	20.42

*Promedio de las tres variables del grupo.

**Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.3.6 Modelo con variaciones sistemáticas y aleatorias y variables ponderadas (SRVMW)

Este último modelo considera tanto las variaciones sistemáticas y aleatorias como la ponderación de las variables. Este modelo representa el modelo más complejo desarrollado en esta fase del estudio. La estimación completa del modelo puede observarse en la Tabla 59.

La consistencia del modelo, al igual que los caos anteriores es adecuado con lo establecido en el apartado 5.1.3.1, considerando para cada variable todas variaciones que le afecten.

La ponderación de las variables sufre una gran variación a lo largo del modelo, puesto que están afectadas por todos los efectos considerados en el mismo, esto repercute en que el análisis del modelo es una tarea compleja. De forma general, la mayoría de parámetros no son considerados aleatorios, en especial aquellos que sufren algún tipo de ponderación. Las mejoras con respecto al resto de modelos más sencillos se discuten de forma más extensa en el apartado 5.1.4.

La ponderación de las variables a afectado a la estructura de los parámetros aleatorios. El parámetro ponderado δn *Tiempo de acceso a la parada es considerado ahora como

aleatorio. El Precio de los billetes se vuelve no significativo de forma independiente, siendo estadísticamente significativo solo cuando se encuentra ponderado e interactuando con los motivos hogar y sanidad. El resto de parámetros mantiene la estructura de los modelos anteriores.

Tabla 59 Resultado del modelo con interacciones, variables ponderadas y parámetros aleatorios (SRVMW)

SRVMW		
F.Log-Verosimilitud	-461.817	
AIC/N	1.327	
Variable	Coef.	Test t
Constante	-5.73758	-11.69
Tiempo de acceso a la parada	0.35827	4.66
Tiempo de espera	0.1807	2.4
Tiempo de viaje	0.26199	2.99
Servicio ofertado (frecuencias)	0.09707	1.35
Fiabilidad del Servicio	0.27713	3.4
Cobertura de las líneas	0.21433	2.6
Forma de conducción	0.40202	5.47
Motivo: trabajo	-1.06449	-2.58
Interacciones		
Edad_25_35* Tiempo de viaje	0.62038	3.11
Trabajador * Servicio nocturno / fin de semana	0.13404	2.29
Motivo: Trabajo * Cobertura de las líneas	0.55626	3.55
Motivo: Estudios* Tiempo de espera	0.16976	1.77
Motivo: Ocio * Tiempo de acceso a la parada	0.19402	2.64
Variables ponderadas		
$\delta n * \delta m$ * Tiempo de viaje	-0.21406	-2.9
δn * Forma de conducción * Viajes > 30	0.45992	2.78
$\delta n * \delta m$ * Precio del billete * Motivo: Casa	0.27849	3.78
$\delta n * \delta m$ * Precio del billete * Motivo: Sanidad	0.61475	4.02
$\delta n * \delta m$ * Tiempo de acceso a la parada * Motivo: Compras	0.56547	3.11
Medias de los parámetros aleatorios		

Facilidad de transbordo	0.25988	3.21
Información*	0.45028	3.89
Confort**	0.94291	7.84
δn * Tiempo de acceso a la parada	-0.15761	-2.27
Parámetros de escala de las distribuciones de los parámetros aleatorios		
Facilidad de transbordo	0.21199	9.61
Información*	0.14178	6.9
Confort**	0.16919	7.67
δn *Tiempo de acceso a la parada	0.11354	3.42
Heterogeneidad en las medias de los parámetros aleatorios		
Estudiante*Facilidad de transbordo	0.16848	2.39
Jubilado*Facilidad de transbordo	0.79904	2.95
Viajes_5_15*Facilidad de transbordos	0.1106	2.63
Jubilado* Confort**	-0.73297	-2.56
Edad_25_35* Confort**	-0.61993	-3.02
Parámetros de límite		
$\mu 1$	2.45654	12.85
$\mu 2$	6.50259	21.4

*Promedio de las tres variables del grupo.

**Promedio entre Confort de los autobuses y Aire acondicionado/ calefacción.

5.1.4 Comparativa de los modelos

A continuación, se muestra la comparativa de los distintos modelos estimados en los subapartados anteriores (Tabla 60). La comparativa se ha realizado basándose en los siguientes indicadores: AIC/N (Akaike, 1973), log-verosimilitud y *Cuenta R^2* (William H Greene and Hensher, 2010; Scott and Freese, 2006). Los valores AIC/N y log-verosimilitud son útiles para comparar distintos modelos, sin embargo, no muestran la capacidad real de predicción de los modelos. En el caso del AIC/N, un modelo es mejor que otro cuanto menor sea el valor de este indicador. En cambio, en el caso de la Log-verosimilitud, un modelo se considera mejor que otro al estar el valor de este indicador más cerca de 0, siendo el valor 0 indicativo de un modelo perfecto. El indicador Count R^2 por el contrario, es más intuitivo y mejor indicador de la capacidad del modelo para representar la realidad. Este indicador se calcula comparando las predicciones correctamente realizadas con el número de observaciones utilizados para estimar el

modelo. De esta forma se calcula el porcentaje de aciertos que tiene el modelo. En el caso de los modelos con parámetros el cálculo de este último indicador es más complejo por lo que se ha realizado la comparativa basándose en los otros dos indicadores.

Tabla 60 Comparativa de los modelos

Modelo	Count R²	Log-Verosimilitud	AIC/N
BM	0.724	-497.064	1.371
SVM	0.743	-467.011	1.331
SVMW	0.750	-464.097	1.323
RBM	N/A	-494.197	1.374
SRVM	N/A	-464.444	1.335
SRVMW	N/A	-461.817	1.327

Cabe mencionar que el modelo inicial estimado en este estudio (BM) es en sí suficientemente acertado para poder asistir la toma de decisiones enfocadas en la mejora de la calidad percibida del servicio. Observando las mejoras generadas en el ajuste de los modelos al incluir los cambios se puede concluir que la mayor aportación reside en la inclusión de la heterogeneidad del usuario mediante interacciones, esto es, al considerar la variabilidad sistemática en los gustos. Este hallazgo se encuentra en consonancia con estudios previos relativos al estudio de la calidad (Bordagaray et al., 2014; dell’Olio et al., 2010), donde se hacía constar la importancia del análisis de la heterogeneidad de los usuarios en un sistema de transporte público. Cuando se consideran parámetros aleatorios se puede observar que el indicador de la Log-verosimilitud mejora a medida que el modelo se vuelve más complejo, mejorando además los valores de los indicadores de sus respectivas parejas de parámetros constantes. Sin embargo, si se observa el valor de AIC/N se puede ver que el indicador muestra un empeoramiento del modelo. Esto es debido a que el valor de AIC/N depende del número de parámetros a estimar incluidos en el modelo en comparación a la capacidad de predicción del mismo. Al incluir las variaciones sistemáticas y aleatorias o las ponderaciones en los modelos, el número de parámetros a estimar crece considerablemente, mientras que el ajuste del modelo no mejora tanto. Por esa razón, lo que indica el valor del AIC/N es que la eficiencia de los modelos con parámetros aleatorios no es buena y que el aumento de parámetros a estimar no se justifica con la mejora del ajuste asociada.

Por ultimo comentar que el modelo que mejor ajuste ha mostrado en este estudio es el modelo más complejo estimado, esto es, el modelo donde se consideran las variaciones sistemáticas y aleatorias en los gustos y la ponderación de las variables mediante la importancia obtenida en el ranking (el modelo SRVMW).

5.1.5 Análisis de los resultados considerando datos faltantes

Una segunda aplicación realizada con los datos obtenidos en la encuesta de satisfacción de 2015 es el análisis considerando datos faltantes. Los resultados mostrados a continuación son los mismos publicados en el segundo artículo que comprende el compendio presentado en esta tesis.

En la Figura 18 se muestra la comparativa realizada entre los resultados de las valoraciones realizadas en la encuesta completa (véase apartado 5.1.2) y las valoraciones calculadas utilizando la base de datos con la información faltante.

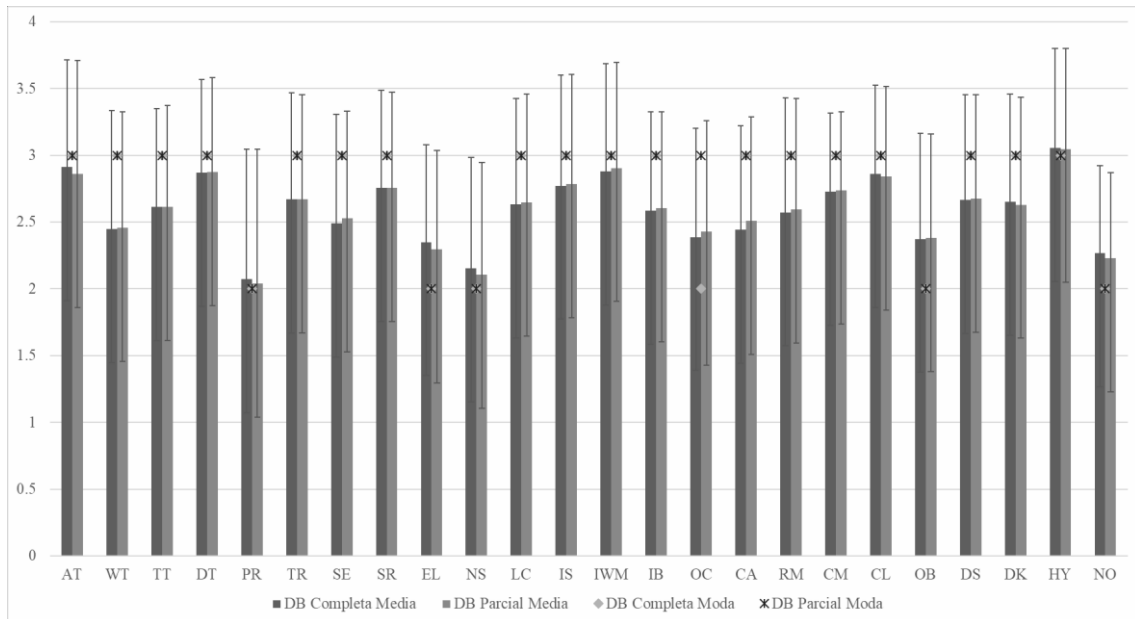


Figura 18 Comparativa de la valoración de los atributos considerando la base de datos completa y con datos faltantes

La comparativa entre las dos bases de datos muestra que, aunque se disponga solo de la mitad de la información, las diferencias en los resultados de la satisfacción de los distintos atributos entre las dos bases de datos son muy bajas. La diferencia entre los promedios es menor al 3% en todos los atributos. La mayor diferencia se observa en la moda del atributo Ocupación (OC), donde la moda cambia de un nivel “Normal” (valor 2) a un nivel “Bueno” (valor 3). Una explicación a este hecho pueda ser que, debido al carácter aleatorio de la eliminación de los datos, es posible que en este atributo concreto se hayan eliminado un valor más alto de valoraciones negativas que positivas. Sin embargo, esta diferencia se ha mostrado únicamente en un atributo por lo que puede considerarse un outlier. La desviación estándar muestra en general un error inferior al 3%, si bien, en el caso del tiempo de acceso a la parada (AT) el error alcanzado es del 6%. Considerando el carácter cualitativo de la encuesta y el limitado error observable entre las dos bases de datos, se puede concluir que los resultados obtenidos mediante ambas bases de datos son similares.

5.1.5.1 Resultados de la modelización con datos faltantes

Se han calculado un total de cuatro modelos para comprobar que es posible establecer unos resultados de modelización equivalentes entre una base de datos completa y otra base de datos con información faltante. Como punto de referencia se ha estimado un modelo denominado BASE, que se ha calculado con la base de datos completa. El resto de modelos se han obtenido partiendo de la base de datos donde el 50% de las valoraciones de los atributos han sido aleatoriamente eliminados. Debido al

requerimiento de los modelos a estimar, ha sido necesario completar la base de datos incompleta. Para completar la información faltante se han seguido los métodos de imputación descritos en el apartado 3.5.3. Una vez completadas las bases de datos se han estimado un modelo por cada método de imputación utilizado, concluyendo con un total de tres modelos cuyos nombres se asocian al método de imputación utilizado para completar la base de datos: Mode, Attribute e Imputacion multiple (MI).

Los atributos escogidos para cada modelo han sido seleccionados siguiendo un proceso de elección paso a paso, hasta que la totalidad de los parámetros incluidos mostraban un signo correcto y un nivel de significatividad estadística aceptable (véase apartado 5.1.3.1). La Tabla 61 muestra los valores de los parámetros estimados, con lo valores t de cada parámetro entre paréntesis. Los parámetros significativos (al menos al 90%, $t > 1,64$) se encuentran resaltados en negrita. Los parámetros significativos son muy similares entre los distintos modelos. El modelo que más se asemeja al modelo BASE en cuanto a la significatividad de los parámetros es el modelo MODE, donde el 79% de los 24 parámetros muestran un nivel de significatividad estadística similar al modelo BASE. El modelo estimado mediante los datos completados con modelos específicos de atributos (ATTRIBUTE) muestra la menor correlación, con un 71% de las variables con un nivel de significatividad similar. El modelo basado en los datos imputados mediante la Imputación Múltiple (MI) queda a medio camino de ambos con una coincidencia del 75%.

Al igual que en los apartados anteriores, debido a que no se recibió ninguna valoración de la Satisfacción General de “Muy Mal”, el valor 0 del modelo representa la agrupación de las respuestas “Muy Mal” y “Mal”. Condicionando la necesidad de estimar solo dos parámetros de límite.

Tabla 61 Resultados de los modelos con datos faltantes estimados

Parámetros	Base	Mode	Attribute	MI ⁶
Constante	-4.37 (-11.22)	-7.81 (-11.89)	-8.47 (-15.25)	-3.81 (-10.82)
Tiempo de acceso a la parada	0.23 (3.71)	0.26 (3.35)	0.35 (4.37)	0.27 (5.05)
Tiempo de espera	0.13 (2.14)	0.18 (2.59)	0.36 (4.73)	0.22 (4.37)
Tiempo de viaje	0.18 (2.47)	0.25 (3.11)	0.39 (4.52)	0.16 (3.37)
Tiempo desde la parada al destino final	0.12 (1.65)	0.24 (2.56)	-	-
Precio del billete	0.07 (1.37)	0.09 (1.34)	0.07 (1.14)	-
Facilidad de transbordo	0.26 (4.2)	0.31 (3.93)	0.36 (4.23)	0.24 (4.07)
Servicio Ofertado (Frecuencias)	0.09 (1.29)	0.29 (3.88)	0.13 (1.53)	0.19 (3.21)

⁶ Promedio de los valores obtenidos en los modelos basados en Imputación Múltiple

Fiabilidad del Servicio	0.19 (2.69)	0.30 (3.49)	0.30 (3.26)	0.02 (1.7)
Líneas especiales	-	-	-	0.1 (2.86)
Servicio nocturno / fin de semana	0.14 (2.38)	-	0.22 (2.78)	0.12 (2.35)
Cobertura de las líneas	0.25 (3.92)	0.32 (4.2)	0.31 (3.91)	0.14 (2.94)
Informacion en las paradas	0.11 (1.78)	0.22 (2.82)	0.19 (2.27)	0.02 (2.44)
Informacion en la web y en el movil	-	0.19 (2.32)	0.27 (3.28)	0.08 (1.64)
Información a bordo	-	-	0.31 (3.56)	-
Nivel de ocupación	-	-	0.18 (2.41)	0.03 (2.85)
Aire acondicionado / calefaccion	0.26 (3.99)	0.16 (2.14)	0.37 (4.52)	0.19 (3.58)
Espacio para personas de movilidad reducida	-	-	0.24 (3.09)	0.04 (4.2)
Confort de los autobuses	0.37 (4.29)	0.30 (2.99)	-	0.42 (5.57)
Limpieza	-	-	-	-
Posibilidad de portar grandes objetos	-	0.16 (2.24)	-	-
Forma de conducción	0.30 (4.55)	0.23 (2.96)	0.25 (3.04)	0.25 (4.28)
Amabilidad del conductor	-	0.14 (1.82)	-	-
Autobuses híbridos	-	-	-	-
Ruido	-	-	0.21 (2.24)	-
Mu(01)	1.73 (17.21)	1.52 (18.02)	2.31 (17.01)	1.73 (17.21)
Mu(02)	4.60 (33.49)	4.09 (37.05)	5.55 (30.2)	4.6 (33.37)
Log-Verosimilitud	-494.86	-675.79	-603.9	-510.26
Cuenta-R ²	0.73	0.63	0.70	0.71
Grados de libertad	17	19	20	15

Dado que la satisfacción general y la satisfacción de los atributos ha sido medida siguiendo la misma escala Likert, la comparativa de los parámetros puede realizarse de forma directa. Se puede considerar que el valor del parámetro ofrece un indicador de la importancia que un atributo tiene para explicar la satisfacción general que lo usuarios tienen con el servicio. El parámetro con el parámetro más alto es el Confort de los autobuses (CM), echo que se repite en los modelos BASE, MODE y MI. El Confort de los autobuses es seguido en importancia por la Forma de conducción (DS), que también puede considerarse un atributo relacionado con la comodidad del viaje. El Precio del billete (PR) muestra un valor del parámetro bajo, lo cual puede interpretarse que el precio

del billete no es importante para la satisfacción de los usuarios. Los atributos relacionados con el servicio (Tiempo de viaje (TT), Tiempo de espera (WT)...) muestran una importancia media y aquellos atributos que son claramente secundarios al servicio básico, como son las Líneas especiales (EL) o el Ruido (NO), han resultado no ser estadísticamente relevantes. Esta tendencia es similar en todos los modelos estimados. Por lo tanto, se puede decir que la satisfacción de los usuarios está altamente condicionada por la comodidad del viaje. Es posible que los usuarios se encuentren adaptado al nivel de servicio actual del sistema, y por ende, apreciar más la comodidad para realizar los trayectos.

A la hora de comparar los distintos modelos, aunque se hayan estimado basándose en datos similares medidos en la misma escala, no es posible hacerlo directamente puesto que los valores de las constantes y de los parámetros límite son diferentes. Por esta razón, se han estandarizado los valores de los parámetros antes de compararlos entre ellos. En la Figura 19 puede observarse la comparativa de los parámetros normalizados de los distintos modelos y se observa una clara correlación entre ellos. Sin entrar en mucho detalle, se puede observar que el modelo MI sigue una tendencia similar al modelo BASE. Lo que quiere decir que la variación de los parámetros es similar, aumentando el valor de los mismos en el modelo MI cuando en el modelo BASE son también altos.

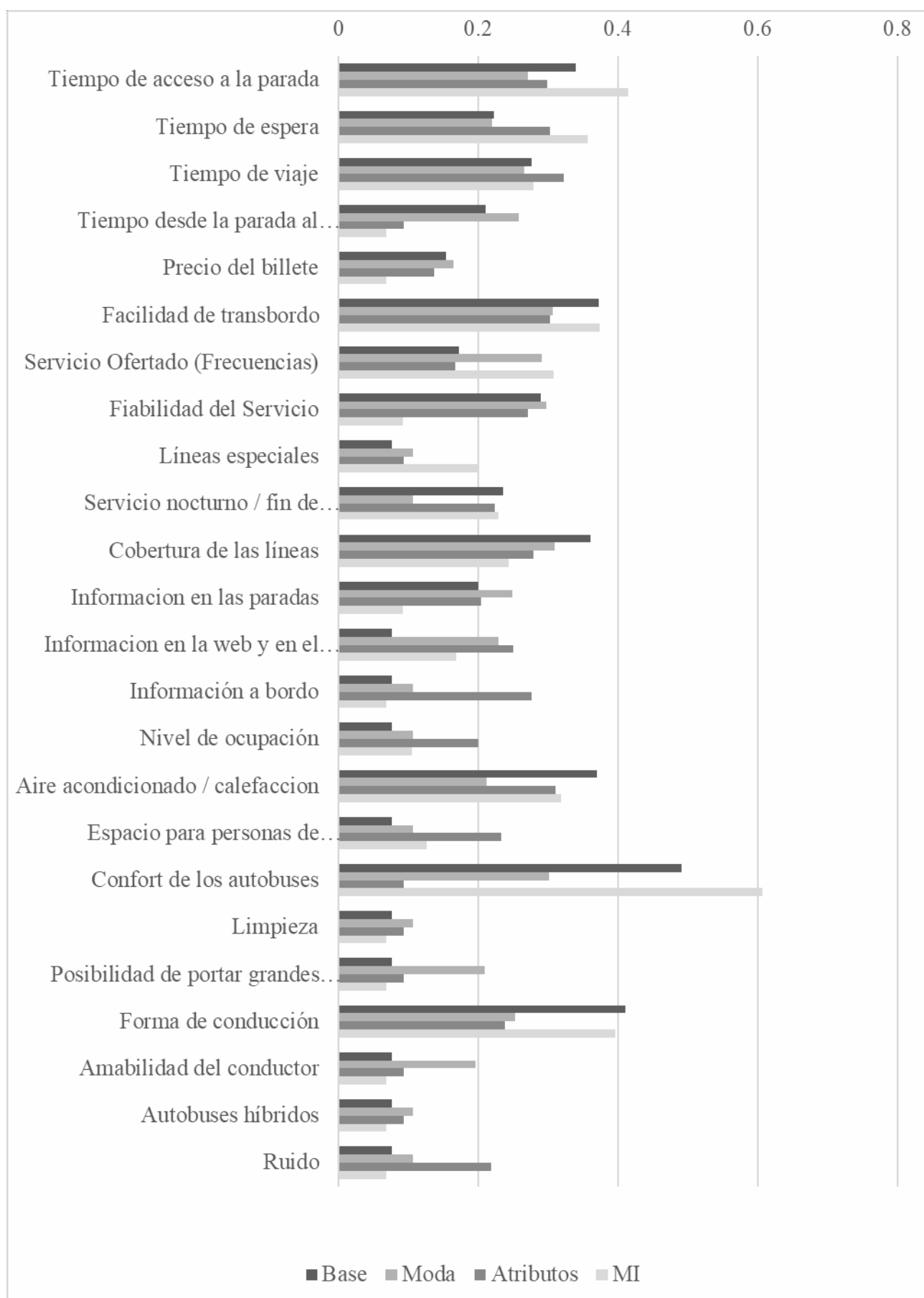


Figura 19 Comparativa de los parámetros estimados en los modelos con datos faltantes.

La correlación no es Buena en todos los caos, en algunos parámetros puede verse cierta discrepancia entre un modelo u otro. Por ejemplo, el Tiempo desde la parada al destino final (ET) y el Precio del billete (PR) no son significativos en el modelo MI pero sí en el modelo BASE. El coeficiente de correlación Pearson calculado entre los modelos muestra una alta correlación. Los valores de los coeficientes de correlación de los distintos modelos comparados con el modelo BASE son 0.95 para MODE, 0.97 para ATTRIBUTTE y 0.99 para MI.

Finalmente, es necesario comparar la capacidad de predicción que muestra cada modelo. Esto puede definirse mediante los indicadores de ajuste como la Log-Verosimilitud o la Cuenta R2. La Tabla 61 muestra los valores de estos indicadores para cada modelo. El mejor ajuste lo muestra el modelo BASE con un valor de Log-verosimilitud de -494.86 y una Cuenta R2 de 0.73. Los demás modelos no alcanzan el ajuste mostrado por el modelo base, sin embargo, las diferencias son pequeñas, con el modelo MI mostrando solo un 1% menos de capacidad de predicción. El modelo MODE muestra el peor ajuste, perdiendo un 10% de capacidad de predicción.

De forma adicional a los indicadores de predicción se ha realizado el test Vuong (Vuong, 1989). Este test permite comparar dos modelos para comprobar que ambos se comportan de forma similar o no. Los resultados del test Vuong se muestran en la Tabla 62. Los valores Z cercanos a 0 significan que dos modelos se comportan de forma similar. Por el contrario, valores Z altos representan modelos con comportamientos distintos. Se establece un límite superior de $Z=1.96$ que considera que dos modelos se comportan de forma diferente a un intervalo de confianza del 95%. La comparativa se debe realizar por pares, comparando el modelo 1 frente al modelo 2. Si el valor Z es positivo quiere decir que el modelo 2 se comporta peor y viceversa. El único modelo basado en imputación que se comporta de forma similar al modelo BASE es el modelo MI con un valor $Z=0.39$. Por lo tanto, se puede considerar que aunque el modelo MI es en inicio peor que el modelo Base su comportamiento estadístico puede considerarse equivalente. Los otros dos modelos muestran valores muy superiores al 1.96, por lo que se puede afirmar fehacientemente que se comportan de forma diferente al modelo BASE.

Tabla 62 Test Vuong para modelos no anidados

Z (Modelo 1 vs Modelo 2)		Modelo 2			
		Base	Mode	Attribute	MI
Modelo 1	Base	0.00	7.77	4.23	0.39
	Mode	-	0.00	-2.29	-7.21
	ATT	-	-	0.00	-3.83
	MI	-	-	-	0.00

5.2 ENCUESTA DE SATISFACCIÓN TUS 2017

5.2.1 Resultados socioeconómicos de la muestra

Esta segunda encuesta de satisfacción se desarrolló durante los meses de octubre y noviembre de 2017. Se encuestaron un total de 4 líneas gestionadas por el operador público de Santander TUS. Se obtuvieron un total de 808 encuestas completas repartidas a lo largo de todo el día.

La Tabla 63 muestra las principales características de los usuarios encuestados. Las mujeres volvieron a estar ligeramente sobre representadas en la encuesta (67% del total), si bien, se observó en los servicios encuestados las mujeres utilizaban más el transporte público que los hombres. Un cuarto de los encuestados fue menor de 25 años, mientras que el resto de grupos de edad se encuentran más equilibrados. En cuanto a la situación laboral, casi la mitad (47%) de los encuestados eran trabajadores, cerca de un cuarto (24%) estudiantes, un 17% jubilados, un 8% desempleados y 5% se dedicaban a las labores del hogar. La mitad de los encuestados disponía de algún otro medio motorizado para realizar el mismo viaje, mientras que solo un 6% estaba dispuesto a realizarlo en bicicleta. Las encuestas realizadas mostraron la importancia que el transporte urbano en Santander tiene para desplazamientos cotidianos, esto es, para los viajes realizados entre el hogar y el trabajo o los estudios. Cerca de la mitad de los de los viajes (46%) tenían como origen el hogar y más del cuarto de los viajes (29%) lo tenían como destino. El trabajo fue el segundo motivo más importante tanto de origen como de destino. La mayoría de los encuestados se trataban de usuarios frecuentes con una baja utilización diaria, más de la mitad (54%) utilizaba el servicio de autobús hasta un máximo de 15 veces por semana. Finalmente, debido al carácter sensible de la pregunta, un 42% de los encuestados decidieron no responder sobre su nivel de ingresos. Los encuestados que si respondieron a esta pregunta tenían principalmente un nivel salarial medio, con un 20% de los encuestado con salarios de entre 900 y 1500 euros y un 17% con salarios entre 1500 y 2500 euros.

Tabla 63 resultados socioeconómicos de la muestra

Genero	Hombre	33%
	Mujer	67%
Edad	< 25	25%
	25 - 34	14%
	35 - 44	15%
	45 - 54	17%
	55 - 64	15%
	65 - 75	11%
	> 75	4%
Situación laboral	Labores del hogar	5%
	Trabajador	47%
	Desempleado	8%
	Estudiante	24%

	Jubilado	17%	
Otros modos de transporte disponibles	Coche (Conduciendo)	35%	
	Coche (acompañante)	12%	
	Bicicleta	6%	
	Moto	3%	
	Otro	44%	
Motivo del viaje		Origen:	Destino:
	Casa	46%	29%
	Trabajo	22%	25%
	Estudios	9%	13%
	Sanidad	4%	5%
	Compras	5%	7%
	Ocio	10%	13%
	Otro	5%	9%
Número de Viajes realizados en bus por semana	< 5	26%	
	5 - 15	54%	
	15 - 30	18%	
	> 30	1%	
Ingresos mensuales	< 900€	7%	
	900€ - 1500€	20%	
	1500€ - 2500€	17%	
	> 2500€	14%	
	Sin respuesta	42%	

5.2.2 Resultados de la valoración de los atributos

La Tabla 64 muestra la satisfacción de los usuarios de los 24 atributos de servicio evaluados en orden descendente, La escala Likert cualitativa de 5 puntos se ha recodificado para el análisis descriptivo y la modelización. Los valores se reparten entre 0 (Muy Mal) y 4 (Muy Bien).

Tabla 64 valoraciones de la satisfacción

Orden	Atributo	Acrónimo	Media	Moda	Desviación estándar
1	Autobuses híbridos	HY	3.24	3	0.77
2	Tiempo de acceso a la parada	AT	2.94	3	0.90
3	Tiempo desde la para al destino final	DT	2.91	3	0.89
4	Limpieza	CL	2.81	3	0.70
5	Facilidad de transbordo	TR	2.79	3	0.90
6	Información en las paradas	IS	2.76	3	0.97
7	Información a bordo	IB	2.73	3	0.90
8	Confort de los buses	CM	2.71	3	0.74
9	Fiabilidad del servicio	SR	2.70	3	0.86

10	Amabilidad del conductor	DK	2.63	3	0.85
11	Calidad de las paradas	ST	2.62	3	0.81
12	Información en aplicación móvil	IM	2.61	3	1.27
13	Cobertura de las líneas	LC	2.60	3	0.83
14	Información en la página web	IW	2.58	3	0.96
15	Espacio para personas de movilidad reducida	RM	2.51	3	0.89
16	Tiempo de espera	WT	2.50	3	0.91
17	Nivel de ocupación	OC	2.50	3	0.87
18	Claridad del mapa de servicios	MD	2.48	3	0.98
19	Tiempo de viaje	TT	2.47	3	0.85
20	Frecuencia de los servicios y horarios	SE	2.44	3	0.97
21	Forma de conducción	DS	2.39	3	0.86
22	Precio del billete	PR	2.33	3	0.94
23	Calefacción / aire acondicionado	CA	2.31	3	0.99
24	Ruido	NO	2.28	2	0.82
	Satisfacción general	OS	2.69	3	0.80

De forma general, los usuarios se encuentran satisfechos con el Servicio, evaluando en promedio con un 2.69 de 4. El atributo con el que los usuarios se encuentran más satisfechos es con el uso de Vehículos Híbridos (HY), siendo este el único atributo superando en promedio el valor 3. Los peores niveles de satisfacción se observaron en el Ruido (NO), el sistema de Aire acondicionado o Calefacción (CA) y el precio (PR).

5.2.3 Modelización mediante modelos de elección discreta

Se estimaron tres modelos utilizando Nlogit v6.0: dos modelos Logit multinomial (MNL) basados en los datos de BW obtenidos de los ejercicios de elección de BW y uno Ordered Logit (OL). La Tabla 65 muestra las estimaciones de los parámetros con los valores de t mostrados entre paréntesis. Los modelos BW muestran el orden de los atributos en términos de niveles de satisfacción/importancia. En los modelos basados en datos BW, el parámetro asociado con el atributo que tiene el nivel más bajo de satisfacción/importancia se fija en 0, lo que permite que todos los demás parámetros sean positivos, esto ayuda a la interpretación de los resultados. Concretamente, el valor de un parámetro identifica su posición en la escala de satisfacción/importancia, siendo un parámetro mayor representativo de un atributo más satisfactorio/importante.

La columna restante muestra los valores de los parámetros del modelo OL. El modelo OL se completa con el término constante y los parámetros de límite (μ), como se explicó en la sección 3.2. La variable dependiente de este modelo es la Satisfacción general (OS) y las variables independientes son todos los atributos evaluados mediante la escala de calificación cualitativa convencional. Los valores de los parámetros en el modelo OL muestran cuánto contribuye cada atributo a explicar la satisfacción general del cliente con el servicio.

Tabla 65 Resultados de los modelos OL y MNL basados en BW

Variable	MNL- BW (Satisfacción)	MNL - BW (Importancia)	Ordered Logit
Autobuses híbridos	1.326 (10.96)	1.083 (8.66)	0.039 (1.87)
Tiempo de acceso a la parada	1.105 (9.34)	1.46 (11.88)	0.151 (2.80)
Tiempo desde la para al destino final	1.054 (8.9)	1.630 (13.1)	0.067 (3.07)
Limpieza	0.794 (6.84)	0.890 (7.36)	0.108 (2.29)
Facilidad de transbordo	0.914 (7.84)	1.161 (9.41)	0.134 (4.24)
Información en las paradas	0.860 (7.27)	1.319 (10.65)	0.155 (2.86)
Información a bordo	0.455 (3.91)	0.249 (2.01)	0.059 (2.68)
Confort de los buses	0.666 (5.71)	1.112 (9.02)	0.063 (2.63)
Fiabilidad del servicio	0.763 (6.62)	2.336 (18.6)	0.270 (4.56)
Amabilidad del conductor	0.480 (4.18)	0.624 (5.07)	0.021 (1.98)
Calidad de las paradas	0.284 (2.41)	0.584 (4.67)	0.154 (2.76)
Información en aplicación móvil	0.585 (4.93)	0.958 (7.68)	0.045 (2.23)
Cobertura de las líneas	0.548 (4.67)	2.235 (17.56)	0.200 (3.59)
Información en la página web	0.494 (4.4)	0 (0)	0.101 (3.35)
Espacio para personas de movilidad reducida	0.406 (3.45)	1.621 (12.89)	0.055 (2.47)
Tiempo de espera	0.375 (3.21)	1.988 (15.76)	0.142 (3.14)
Nivel de ocupación	0.209 (1.81)	1.478 (12.03)	0.145 (3.28)
Claridad del mapa de servicios	0.114 (0.98)	1.118 (8.97)	0.180 (3.25)
Tiempo de viaje	0.434 (3.66)	2.239 (17.29)	0.080 (3.71)
Frecuencia de los servicios y horarios	0.305 (2.62)	2.260 (17.82)	0.298 (5.11)
Forma de conducción	0.075 (0.65)	1.474 (11.8)	0.078 (2.41)
Precio del billete	0 (0)	1.860 (15.03)	0 (0)
Calefacción / aire acondicionado	0.050 (0.43)	0.430 (3.49)	0.062 (2.86)
Ruido	0.069 (0.6)	0.472 (3.87)	0.030 (2.66)

Constante	-	-	-1.305 (-2.88)
Mu(01)	-	-	1.427 (11.07)
Mu(02)	-	-	3.628 (33.61)
Mu(03)	-	-	6.997 (46.47)
Log-Verosimilitud	-5817.674	-5373.256	-818.697
AIC/N	4.819	4.452	2.072
McFadden PseudoR ²	-	-	0.11

Como se indica en el modelo de satisfacción BW, los usuarios están muy satisfechos con la política ambiental de la empresa al utilizar autobuses híbridos (HY). Además, los usuarios también están muy satisfechos con los Tiempos de acceso a las paradas (AT) y los Tiempo desde la para al destino final (DT), lo que significa que los usuarios de los autobuses en Santander ven una buena cobertura espacial de las paradas.

En cuanto a los atributos menos satisfactorios, el Precio de los billetes (PR) se encuentra como el menos satisfactorio. Las tarifas actuales de los billetes varían según el tipo de usuario y el sistema de pago utilizado. Sin embargo, el precio del servicio no es superior al de otros servicios de transporte público de ciudades cercanas. El resultado puede sugerir la existencia de un comportamiento de respuesta estratégica, en el sentido de que los encuestados votaron estratégicamente de forma negativa sobre su satisfacción con las tarifas de transporte para reducir la posibilidad de que los operadores aumenten las tarifas en el futuro. Este comportamiento también se observó en la anterior encuesta presentada en los apartados anteriores. Las características ambientales como el Ruido (NO) y los sistemas de Aire acondicionado / calefacción (CA) también muestran bajos niveles de satisfacción, así como varios atributos relacionados con el confort durante el viaje: la Forma de conducción (DS) y el nivel de Ocupación (OC). Analizando los canales de información existentes en el servicio, se observa que los usuarios están muy satisfechos con la información ofrecida en las paradas (IS), y algo menos con la información disponible en las aplicaciones móviles (IM). Si bien se muestra que para el resto de las fuentes de información (información a bordo de los autobuses (IB) y en el sitio web (IW)) los usuarios no están satisfechos.

El modelo basado en respuestas BW de importancia muestra que los atributos más importantes son los que están directamente relacionados con las características básicas del servicio, como la Fiabilidad del servicio (SR), la Frecuencia (SE), el Tiempo de viaje a bordo (TT) y la Cobertura de las líneas (LC). Por el contrario, los atributos menos importantes son el nivel de Ruido (NO), los sistemas de Aire acondicionado / calefacción (CA), la Información a bordo (IB) e Información en la página web (IW).

Según el modelo OL, los atributos que muestran un alto valor de los parámetros son la Frecuencia de servicio (SE), la Fiabilidad del servicio (SR) y la Cobertura de las líneas (LC). Por el contrario, la Amabilidad del conductor (DK), el nivel de Ruido (NO) y el uso de tecnologías híbridas (HY) son los que tienen las estimaciones de parámetros más

bajas. Los parámetros de límite muestran una no linealidad en los diferentes puntos de valoración, lo que significa que, desde la perspectiva del usuario, se requieren diferentes niveles de esfuerzo para mejorar el servicio en un punto de satisfacción al siguiente, como por ejemplo de muy malo a malo o de bueno a muy bueno.

Con los datos disponibles se quiso analizar si existe alguna conexión entre los modelos derivados del ejercicio Best-Worst y los resultados obtenidos de la modelización del Ordered Logit y los promedios de los niveles de satisfacción (sección 5.2.2). La Tabla 66 muestra la correlación en los parámetros obtenidos mediante los dos enfoques. La correlación entre los promedios de las puntuaciones de satisfacción tradicional para cada atributo y el modelo MNL de satisfacción basado BW es casi perfecta, con un coeficiente de correlación de 0,95. El modelo OL muestra una correlación considerable ($r = 0,486$) con el modelo MNL basado en las respuestas BW de importancia.

Tabla 66 Correlación entre los parámetros de los modelos OL y MNL basados en BW

	BW (Importancia)	BW (Satisfacción)	Ordered Logit	Puntuación de satisfacción
BW (Importancia)	1			
BW (Satisfacción)	0.083	1		
Ordered Logit	0.486	0.074	1	
Puntuación de satisfacción	-0.056	0.946	0.016	1

Una investigación más profunda de estas fuertes correlaciones se presenta en las Figuras Figura 20 y Figura 21. Los valores de los parámetros difieren en escala de un modelo a otro, y por lo tanto una comparación directa de las estimaciones de los parámetros no muestra la verdadera correlación entre los mismos. Por esta razón, para facilitar la comparativa se muestran los valores normalizados. La primera comparativa (Figura 20) muestra la relación entre los dos valores que representan la satisfacción. Por un lado, se encuentran los resultados promedio de las evaluaciones convencionales obtenidas mediante la escala cualitativa, que a efectos de análisis se representan con puntuaciones comprendidas entre 0 y 4. Por otro lado, los parámetros estimados en el modelo MNL basado en los datos de satisfacción BW. Como puede verse en la Figura 20, la correlación entre estos dos valores es considerablemente alta. La mayoría de los atributos muestran una tendencia similar para ambos casos, aunque hay algunas excepciones como la Información en la aplicación móvil (IM), la Cobertura de las líneas (LC), la Información a bordo (IB) y la Calidad de las paradas (ST). Por lo tanto, ambos métodos conducen a los mismos resultados, lo que respalda la hipótesis de que el método BW puede reemplazar el índice de satisfacción convencional.

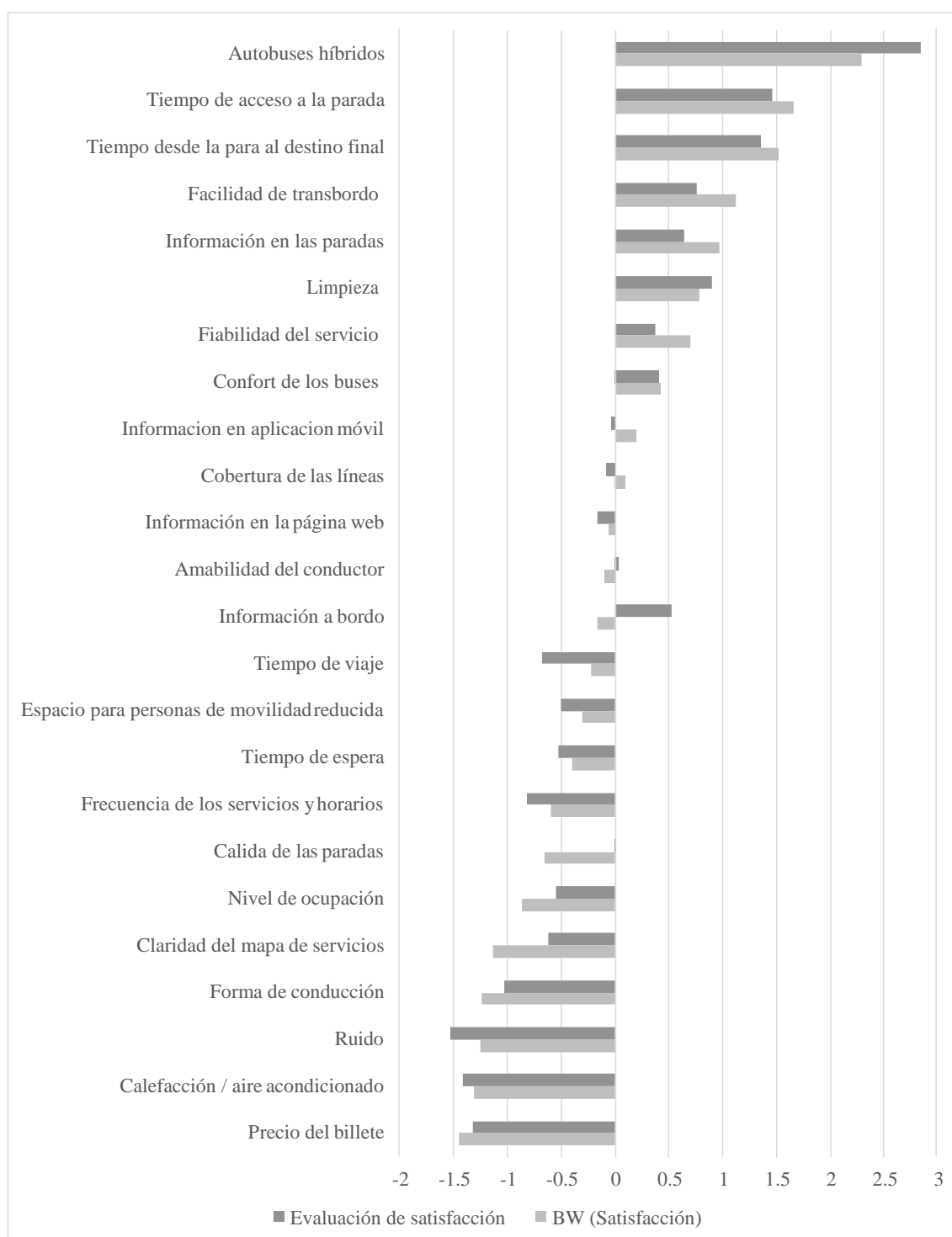


Figura 20 Comparación entre la puntuación de satisfacción y los parámetros del modelo MNL basado en datos BW de satisfacción.

Para la comparación de la importancia, se han seleccionado los modelos OL y MNL basados en las respuestas de importancia BW. El valor que tiene un parámetro en un modelo OL puede asociarse con el peso que tiene este a la hora de explicar la variable dependiente. En otras palabras, el parámetro explica la contribución de cada atributo a la satisfacción global, esto es, la importancia implícita. El análisis de estos dos conjuntos de parámetros, presentados en la Figura 21, muestra un nivel de correlación más bajo que en el caso de la satisfacción. Aunque los niveles de importancia de los atributos más

importantes muestran una tendencia similar en ambos modelos, el resto de los atributos muestran una correlación muy pequeña. En consecuencia, aunque ambos modelos representan un cierto nivel de importancia de las variables, estos valores no son los mismos y, por lo tanto, representan conceptos de importancia diferentes.

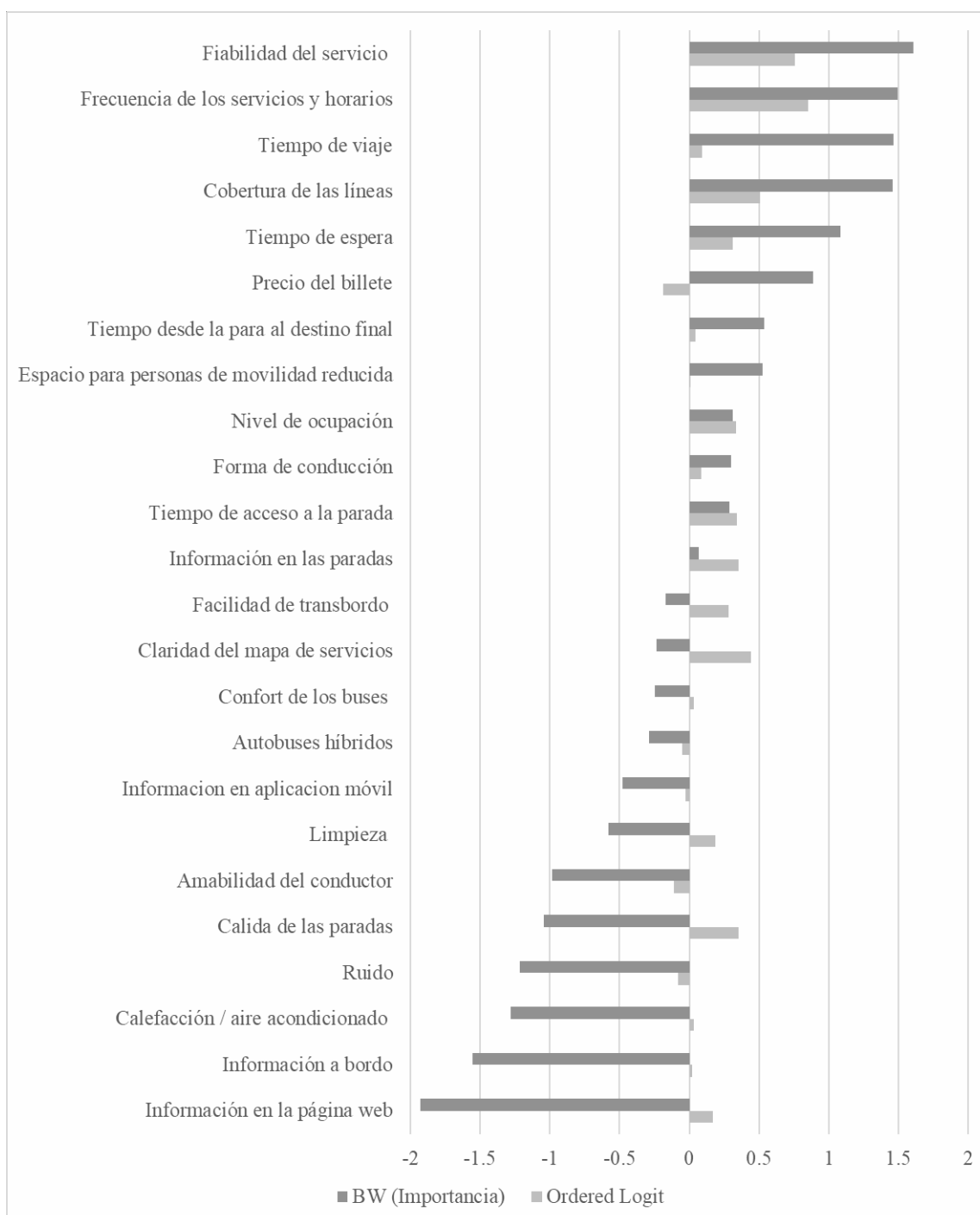


Figura 21 Comparación de lo parámetros del modelo OL y el MNL basado en datos BW de importancia.

5.2.4 Análisis de Importancia-Valoración (IPA: Importance-Performance Analysis)

El análisis de importancia-valoración (IPA) (Martilla and James, 1977) es una herramienta de decisión ampliamente utilizada. La base de este método consiste en cruzar ambos resultados (nivel de satisfacción e importancia) en el mismo gráfico. Se definen cuatro cuadrantes, cada uno de ellos con un nivel diferente de prioridad de mejora de los atributos localizados en los mismos. Los cuatro cuadrantes se identifican típicamente como "mantener el buen trabajo" (Q1 - "Importante y satisfecho"), "posible derroche de recursos" (Q2 - "No importante y satisfecho"), "baja prioridad" (Q3 - "No importante y no satisfecho") y "concentrarse aquí" (Q4 - "Importante y no satisfecho") (Sever, 2015). Los atributos en el Q1 se consideran las fortalezas del servicio, atributos que están funcionando bien y donde las inversiones deben mantenerse para mantener el nivel de satisfacción. Los atributos en el Q2 contienen atributos que no son importantes para los usuarios pero que aun así tienen un nivel de calidad alto, lo que puede indicar que existe un posible derroche de recursos utilizados en estos atributos. Los atributos en el Q3 son los que tienen el menor nivel de prioridad para la inversión. Finalmente, el Q4 muestra las principales prioridades de mejora del servicio, atributos que son importantes para los usuarios pero que no están rindiendo lo suficientemente bien como para satisfacer a los clientes.

El método IPA se utilizó recientemente en la literatura para comparar la importancia explícita e implícita de los servicios de transporte público (Cao y Cao, 2017). La importancia explícita se obtiene preguntando al encuestado de forma directa, ya sea mediante un método de clasificación convencional o algunos otros métodos, mientras que la importancia implícita se deriva de un modelo, por ejemplo, de un modelo de datos ordenados. En el estudio mencionado se comprobó que las prioridades de mejora de los servicios basadas en la importancia explícita son diferentes de las basadas en la importancia implícita. Con el fin de verificar esta importante conclusión, en la Figura 22 se presenta el análisis de Importancia-Valoración (IPA) en el que se comparan la importancia explícita y la satisfacción obtenidas con el modelo BW con las obtenidas con el método convencional (puntuación y modelización mediante datos ordenados). Los niveles de satisfacción se representan en el eje horizontal, mientras que los niveles de importancia se muestran en el eje vertical. Las estrellas muestran las posiciones de los atributos de servicio según los datos y resultados de la modelización de BW. Los cuadrados muestran lo mismo, pero utilizando la clasificación convencional como satisfacción y los parámetros estimados mediante el modelo OL como importancia. Los valores están normalizados a efectos de presentación.

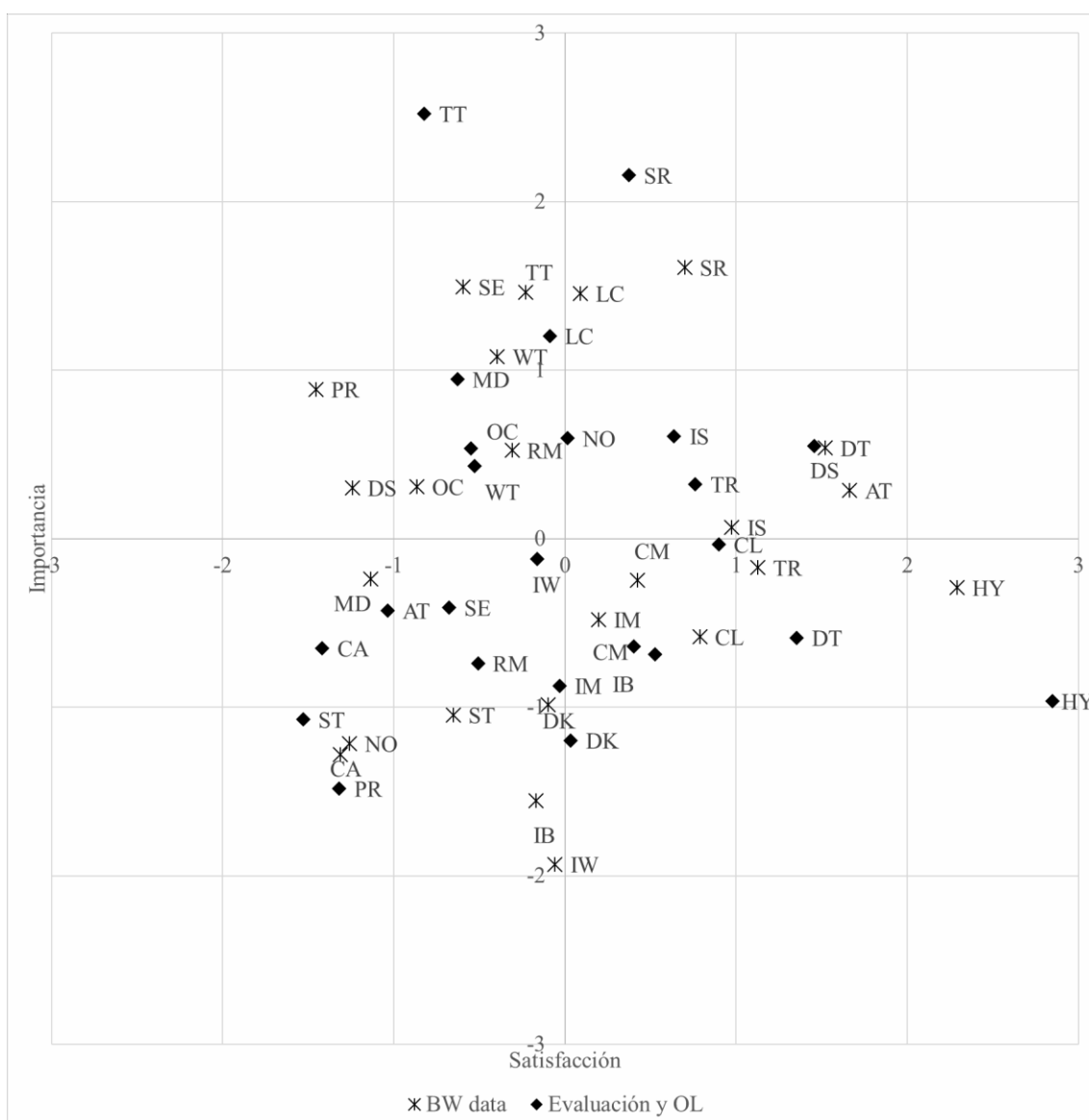


Figura 22 Análisis IPA de la encuesta de satisfacción de 2017

Cerca del 55% de los atributos de servicio se posicionan en un cuadrante distinto cuando se utiliza el método convencional frente al método BW. Para algunos atributos, como la Cobertura de líneas (LC) y la Información en aplicación móvil (IM), los cuadrantes son diferentes pero las posiciones están muy cerca unas de las otras. Sin embargo, en el caso de otros atributos, como la Precio del billete (PR), las diferencias son bastante notables, ya que el método BW identifica la tarifa como una variable importante mientras que el modelo OL estimado sugiere lo contrario (PR no es importante).

Las diferencias entre los dos métodos son mucho mayores en importancia que en satisfacción. De hecho, los valores de satisfacción son bastante similares entre los dos métodos, según las posiciones relativas al eje horizontal. La mayoría de los atributos se colocan en un valor de eje horizontal muy cercano, lo que significa un nivel de satisfacción similar. Por consiguiente, las diferencias de posición en los cuadrantes son el resultado de los diferentes niveles de importancia obtenidos por los métodos alternativos. Dado que la importancia de los atributos derivada del modelo de BW es explícita (ya que los usuarios eligen explícitamente los atributos más y menos importantes de un conjunto

de atributos) y que la del modelo OL es implícita (derivada de las estimaciones de los parámetros de un modelo de satisfacción general), apoyamos el argumento de Beck and Rose (2016) de que la escala de BW es mejor que la clasificación convencional para definir la importancia de los atributos. Además, los resultados obtenidos con el método BW para definir la importancia se encuentra más en línea con la literatura existente.

Utilizando los resultados de los métodos de BW, un cambio en la actual estructura de tarifas puede resultar en un mayor nivel de satisfacción ya que PR es considerado un atributo muy importante, además de ser el atributo con el menor nivel de satisfacción. Sin embargo, los resultados del modelo de datos ordenado OL, sugieren que cualquier cambio en la política de tarifas no generaría un aumento de la satisfacción general del servicio. Esta evidencia contradictoria se aplica también a la frecuencia del servicio en la que el método BW sugiere que mejorar la frecuencia del servicio mejoraría la satisfacción del cliente mientras que el método OL identifica la frecuencia del servicio como una prioridad baja. Estas diferencias sugieren que el concepto de importancia puede diferir entre los dos métodos de modelización. En la siguiente sección se investiga más a fondo este tema.

5.2.5 Modelos de regresión entre los distintos métodos de modelización

Una cuestión pendiente en relación a la utilización del método de encuesta de BW en el estudio de la satisfacción de los usuarios, es si sería posible obtener el nivel de satisfacción de los distintos atributos a partir de los resultados BW. En esta sección se aborda esta cuestión realizando un análisis de regresión para estimar el nivel de satisfacción promedio de cada atributo basado en la Satisfacción general del servicio y los parámetros del modelo de BW. Se han estimado un total de dos modelos: uno para el nivel de satisfacción y otro para el nivel de importancia de cada atributo.

El modelo de satisfacción de BW que se muestra en la sección 5.2.3 representa el nivel de satisfacción relativo de los muchos atributos que juntos definen el servicio completo. Por lo tanto, estos parámetros permiten establecer el nivel de satisfacción que tiene un atributo con respecto a otro que se utiliza como referencia. Por lo tanto, el desafío consiste en convertir este nivel de satisfacción relativo en un nivel de satisfacción promedio para cada atributo. Similar al resultado obtenido mediante las valoraciones realizados en las encuestas de satisfacción convencionales. Se propone como forma de complementar la información ofrecida por el ranking BW, el añadir un término constante al modelo de regresión. De esta forma, se consigue convertir de forma efectiva los parámetros obtenidos en el modelo MNL basa en BW en valoraciones promedio que muestren el nivel de satisfacción real de cada atributo sin necesidad de establecer un atributo base. Más concretamente, es el valor de la Satisfacción general la que actúa como término constante, ya que permanece constante para toda la muestra. El modelo de regresión se especifica en la ecuación (7), en la que la variable dependiente es el nivel de satisfacción de cada atributo (el promedio mostrado en la Tabla 64) y las variables independientes son los la Satisfacción general (OS) y los parámetros del modelo MNL de satisfacción basado en BW (Tabla 65).

$$\text{Nivel de satisfacción}_i = \text{Satisfacción general} + e^{\delta_k^{\text{BW(Satisfacción)}}} \quad (62)$$

Tabla 67 Modelo de regresión para el nivel de satisfacción de los atributos

	Parámetro	Valor t
Satisfacción general	0.774	52.017
exp[BW(Satisfacción)]	0.301	14.397
R ²	0.999	
R ² Ajustado	0.954	

En la Tabla 67 se muestra el resultado de la estimación, lo que confirma que es posible estimar con precisión el nivel de satisfacción promedio de cada uno de los atributos a partir de los datos de BW. El modelo muestra un R² de 0,999, lo que indica que casi el 100% de la variación de la satisfacción específica de cada atributo se explica por el nivel de OS y los parámetros del modelo de BW (es decir, la satisfacción relatividad entre los diferentes atributos). El resultado del modelo de regresión sugiere que las encuestas de valoración convencionales pueden sustituirse por encuestas de basados en BW para medir la satisfacción de los atributos.

Los parámetros asociados a un modelo de datos ordenados definen hasta cierto nivel la importancia implícita de los diferentes variables que lo componen. Como se muestra en la sección 5.2.3, la importancia implícita del modelo OL y la importancia explícita derivada del modelo MNL basada en BW están correlacionadas, aunque la correlación no es tan consistente como en el caso de la satisfacción. En la Tabla 68 se muestra el resultado del modelo de regresión definido en la ecuación (8).

$$\beta^{OL} = \text{Nivel de satisfacción}_i + e^{\phi_k^{BW(\text{Importancia})}} \quad (63)$$

Tabla 68 Modelo de regresión para la importancia de los atributos

	Parámetro	Valor t
Nivel de satisfacción	0.017	1.914
Exp[BW(Importancia)]	0.015	3.282
R ²	0.788	
R ² Ajustado	0.733	

La bondad de los indicadores de ajuste muestra que los parámetros del modelo OL pueden ser estimados utilizando datos BW hasta cierto punto. Esto significa que hay una diferencia en la forma en que se capta la importancia en cada método. En el método BW, la importancia se considera explícita mientras que en el OL es implícita, por lo tanto, se puede concluir que están relacionadas, pero no son iguales. Los parámetros de un modelo de OL agrupan tanto la importancia explícita como la satisfacción. En otras palabras, los parámetros del modelo de OL representan no sólo la importancia que cada atributo tien

dentro del sistema de forma aislada, sino que también considera la satisfacción con respecto a todo el sistema. Esto explica por qué hay tanta diferencia en el análisis del IPA que se muestra en la sección 5.2.4, ya que el modelo OL va más allá de la simple definición de un nivel de importancia de los diferentes atributos.

5.3 ENCUESTA SOBRE LA IMPORTANCIA EN EL FERROCARRIL

5.3.1 Análisis cualitativo

Antes de realizar la encuesta que permitiese definir los atributos más relevantes en un servicio de transporte ferroviario, se desarrolló un trabajo cualitativo enfocado en establecer las características principales de este tipo de servicios. Como resultado principal del proceso consultivo seguido en esta fase, se obtuvo un DAFO que capturaba los distintos aspectos del transporte ferroviario. Los resultados del análisis DAFO consideran de forma conjunta las entrevistas a expertos y los grupos focales (FGs) desarrollados con distintos agentes involucrados en el transporte ferroviario. El DAFO se muestra en la Tabla 69. El análisis DAFO permite obtener un diagnóstico preliminar sobre la situación del sector ferroviario en función de los agentes implicados. La información que se muestra en el DAFO representa el aspecto principal en el que los diferentes agentes han consensuado las diferentes características del sector ferroviario. Esta información se obtuvo mediante el análisis de las transcripciones de los FGs y las entrevistas en profundidad.

Tabla 69 Análisis DAFO (Debilidades, Amenazas, Fortalezas y Oportunidades)

Fortalezas	Debilidades
<ul style="list-style-type: none"> • Buena información, accesibilidad y servicios en estaciones principales. Tanto fuera como dentro de las estaciones. • Tarifas e información tarifaria adecuada. Posibilidad de comprar billetes online y en las estaciones. • Buena accesibilidad para personas de movilidad reducida en estaciones principales. • Localización estratégica de algunas estaciones • Mayor confort que otros medios de transporte. • Buenos servicios a bordo (cafetería, Wi-fi, música, video, vagones silencio, etc.) • Niveles de fiabilidad altos en trayectos de media y larga distancia. 	<ul style="list-style-type: none"> • Dificultad para realizar los transbordos (especialmente para personas mayores) debido a la falta de información. • Servicios de cercanías con frecuencias bajas, poca fiabilidad y prácticamente ningún Servicio adicional. • Coste adicional por la compra online. Mal funcionamiento de la web. Tarifas muy rígidas y con pocos descuentos para los distintos tipos de usuario (familia números, grupos, etc). • Falta de servicios en estaciones no principales, especialmente para personas con movilidad reducida. • Problemas de infraestructura con trenes demasiado largos para las estaciones no principales. Anchos de vía distintos. • Mala conexión geográfica. • Uso repetido del contenido multimedia (música, películas, etc.).

<ul style="list-style-type: none"> • Altos niveles de seguridad contra incidencias y robos. • Buena seguridad en situaciones climáticas adversas. 	<ul style="list-style-type: none"> • Falta de confort en algunos asientos. • Restricciones para viajar con animales o equipaje voluminoso. • Demasiados controles de equipaje
Oportunidades	Amenazas
<ul style="list-style-type: none"> • Posibilidad de introducir servicios de bajo coste. • Financiación pública para compensar parte de los gastos de operación para permitir precios más competitivos. • Posibilidad de realizar controles de equipaje solo en situaciones de alta demanda. • Mejora del Sistema de compra de billetes utilizando tecnologías modernas. • Posibilidad de que las autoridades locales mejoren la accesibilidad a las estaciones mediante transporte público. • Inversiones futuras en el desdoblamiento de las líneas para cercanías. • Tecnologías enfocadas en los andenes para subir al tren sin esperas. • Reducción de la necesidad de mantenimiento debido a la automatización y el mantenimiento predictivo. • Desarrollo de sistemas de realidad aumentada para ofrecer una mejor información en las estaciones. • Nuevos sistemas de predicción que permitan desarrollar servicios a demanda y reaccionar de forma rápida a retrasos e incidentes. 	<ul style="list-style-type: none"> • Página web difícil de usar. La venta automática es una barrera para ciertos sectores de la población. • Localización de las estaciones fuera de la ciudad. • Poco espacio de aparcamiento en las estaciones. Aparcamiento ilegal en ciertas estaciones. • Baja frecuencia comparado con el autobús. • Falta de mantenimiento de la infraestructura. Derivado de una dejadez en la mejora de los servicios. • Baja frecuencia comparado con el autobús. • Alto coste de la infraestructura nueva • Atención al cliente deshumanizado y automatización excesiva. • Baja confianza de los usuarios debido a incidentes mal gestionados. • Mala planificación de las conexiones entre países. • Nuevos competidores para viajes de larga distancia (por ejemplo el Hyperloop).

Entre las fortalezas del sistema destacan los aspectos relacionados con la información, la accesibilidad o los servicios auxiliares y la comodidad, especialmente los relacionados con las estaciones principales y los servicios. Las debilidades, en cambio, están más relacionadas con el sistema tarifario y las deficiencias en ciertos servicios, como las bajas frecuencias, la falta de servicios en las estaciones no principales y la creciente obsolescencia en las líneas que no están adaptadas a la alta velocidad.

Las oportunidades que se mencionan en el sistema ferroviario son la posibilidad de realizar mejoras tecnológicas en áreas como los sistemas de adquisición de billetes, los sistemas de acceso de los usuarios a los trenes, los servicios de realidad aumentada para mejorar la información o los nuevos sistemas que permitan prevenir retrasos o incidentes.

También se mencionó la posibilidad de introducir mejoras en el sistema tarifario, como nuevos servicios de bajo coste, mejoras en la intermodalidad a través de los aparcamientos y el transporte público, y servicios de a bordo más especializados.

Dentro de las amenazas se encuentra la competencia con otros modos de transporte actuales como el autobús, dadas sus mayores frecuencias, y una posible competencia futura con nuevos modos de transporte basados en la tecnología de vacío (es decir, modos similares al Hyperloop). También se mencionó el uso excesivo de la tecnología como una amenaza en caso de dar lugar a un tratamiento excesivamente impersonal a los clientes. Por último, la falta de interoperabilidad entre los sistemas ferroviarios de diferentes países también se consideró una amenaza relevante.

5.3.2 Resultados socioeconómicos de la muestra

La encuesta final se llevó a cabo en mayo de 2017, obteniéndose 316 respuestas completas de un total de 609 participantes, es decir, el 52% de las encuestas fueron completadas. La encuesta se elaboró en una plataforma en línea y se difundió a través de diferentes redes sociales, en las que se invitó a participar a personas de toda Europa. No se ofreció ningún incentivo para participar en la encuesta. Se obtuvieron respuestas de más de 24 países en relación con más de 180 ciudades. Los encuestados resultaron ser en su mayoría hombres (79%) que residían en ciudades de tamaño medio (30%), de edad joven, con una mayoría entre 25 y 34 años. Casi todos los encuestados tenían carnet de conducir, pero no todos (70% de los que tenían carnet) disponían de vehículo propio. Además, los encuestados eran principalmente empleados (55%) con un salario medio. En cuanto al uso del ferrocarril, un gran número de los encuestados lo utilizaba al menos unas pocas veces al mes y la mayoría de ellos tenían una estación cerca de su casa o lugar de residencia. El desglose completo de los datos se muestra en la Tabla 70. A pesar de que la muestra no es representativa de la población europea, las observaciones se ponderaron teniendo en cuenta los estratos de sexo y edad para asegurar su validez. De esta manera, la muestra se adaptó para representar la estructura de la población de la Unión Europea: 49% de hombres y 51% de mujeres; 26% de la población menor de 25 años, 42% entre 25 y 54 años, 13% entre 55 y 64 años, y 19% mayor de 65 años (Central Intelligence Agency, 2017).

Tabla 70 Resultados socioeconómicos de la encuesta de importancia ferroviaria

Género	Hombre	79%
	Mujer	21%
Tamaño de ciudad	Áreas rurales (pob < 25,000)	17%
	Ciudades pequeñas (pob 25,000–150,000)	20%
	Ciudades medianas (pob 150,000–500,000)	30%
	Ciudades grandes (pob 500,000–1,500,000)	17%
	Metrópolis (pob > 1,500,000)	16%
Edad	<25	6%
	25–29	21%
	30–34	15%
	35–39	13%

	40–44	11%
	45–49	11%
	50–54	7%
	55–59	6%
	60–65	4%
	>65	6%
Carnet de conducir	Disponen de carnet de conducir	96%
	Disponen de vehículo propio	70%
Sufren de alguna discapacidad o movilidad reducida		3%
Situación laboral	Trabajador	55%
	Autónomo	9%
	Funcionario	17%
	Estudiante	12%
	Jubilado	5%
	Desempleado	2%
	Otro	0%
Nivel de ingreso	<900 €	10%
	900–1500 €	10%
	1500–2500 €	26%
	2500–3500 €	13%
	3500–4500 €	6%
	>4500 €	10%
	Sin respuesta	25%
Uso del tren	Nunca / ocasionalmente (0–6 veces al año)	38%
	Semanalmente (1–2 veces por semana)	31%
	Frecuentemente / a diario (todos o casi todos los días)	31%
Distancia a la estación de tren más cercana	<1 km (5–10 min caminando)	30%
	1–3 km (10–30 min caminando)	39%
	3–6 km (30–60 min caminando)	19%
	>6 km (más de 1h caminando)	12%

5.3.3 Resultados de la modelización sobre la importancia de los atributos en el ferrocarril

Con los datos BW recogidos, se estimó un modelo Logit Multinomial (MNL), que permite establecer el orden de importancia de los atributos en base a los parámetros calculados. Cuanto más alto es el parámetro de un atributo, mayor es la importancia que los encuestados le han dado. Para que el modelo sea estimable, es necesario establecer un atributo como base, al que se le da el valor del parámetro de 0. En este caso, se eligió el atributo OBP (Posibilidad de llevar animales domésticos a bordo) dado que se consideró el menos importante. Por lo tanto, los coeficientes de todos los demás atributos tendrán valores positivos. En la Tabla 71 se muestran los parámetros estimados.

Tabla 71 Modelo MNL basado en los datos Best-Worst

Código	Atributo	Parámetro	Valor z
1	IA	1.639	8.33
2	AT	2.859	13.41
3	TFT	3.135	15.28
4	TVT	4.216	18.8
5	COM	3.244	15.2
6	CP	1.814	8.85
7	BP	1.672	8.17
8	UT	2.850	14.16
9	IUT	2.406	11.53
10	NUM	4.039	18.7
11	SERV	2.629	12.38
12	OBI	3.068	15.13
13	ENV	1.439	6.88
14	PR	4.076	18.55
15	SEG	0.294	1.58
16	LAN	1.690	8.25
17	TICK	2.645	12.5
18	SOB	2.516	12.42
19	SPO	2.030	10.03
20	OBB	1.703	8.46
21	OBL	1.705	8.63
22	ACC	2.179	10.71
23	RE	1.803	9.27
24	PUR	3.010	14.59
25	ST	1.997	9.96
26	CONU	3.405	16.54

27	CONI	3.817	18.14
28	LOC	3.359	16.25
29	IAD	0.266	1.22
Log-likelihood		-3096.43	
AIC/N		3.956	

5.3.4 Variación de la importancia de acuerdo al tipo de usuario.

Las variables socioeconómicas obtenidas en la encuesta han permitido realizar un análisis considerando diferentes tipos de usuarios. Los resultados más interesantes son los relacionados con los niveles de uso del ferrocarril, ya que permiten diferenciar la importancia dada a los atributos por los usuarios y no usuarios del ferrocarril. Para la modelización se han codificado los diferentes niveles de uso mediante la técnica de codificación de efectos (effect coding). A diferencia de la codificación dummy, en este tipo de codificación el grupo de referencia se identifica con un valor -1. Por lo tanto, el término constante obtenido en el modelo representa la gran media de la importancia de cada atributo para la muestra en su conjunto. Así pues, los coeficientes asociados a cada nivel muestran la desviación de ese grupo con respecto a la media. La ventaja de este tipo de codificación es que es posible calcular el coeficiente asociado al grupo de referencia. El valor de este coeficiente se calcula como la suma de los valores negativos de los coeficientes del resto de los grupos. Las variables codificadas de efecto se han introducido en el modelo como interacciones con cada atributo. Se han definido como grupo base el conjunto de usuarios ocasionales/no usuarios. En la Tabla 72 se muestra el modelo estimado, en el que se han omitido las interacciones que han resultado no ser estadísticamente significativas.

Tabla 72 Modelo MNL con interacciones por frecuencia de uso del ferrocarril

Atributo	Constante	Valor z	Frecuente/a diario	Valor z	Semanalmente	Valor z	Nunca/Ocas.
IA	1.664	8.40	-	-	-	-	-
AT	2.908	13.49	-	-	0.466	2.81	-0.466
TFT	3.176	15.38	-	-	-	-	-
TVT	4.307	18.86	-	-	0.363	1.99	-0.363
COM	3.282	15.28	-	-	-	-	-
CP	1.777	8.54	-0.567	-3.66	-	-	0.567
BP	1.723	8.36	-	-	0.427	2.78	-0.427
UT	2.893	14.27	-	-	-	-	-
IUT	2.449	11.64	-	-	-	-	-
NUM	4.150	18.82	0.595	3.24	-	-	-0.595
SERV	2.653	12.42	-	-	-	-	-
OBI	3.127	15.28	-	-	-	-	-
ENV	1.490	7.07	-	-	0.282	1.63	-0.282
PR	4.135	18.52	-0.275	-1.49	-	-	0.275

SEG	0.320	1.71	-	-	0.339	2.16	-0.339
LAN	1.736	8.42	-	-	-	-	-
TICK	2.682	12.58	-	-	-	-	-
SOB	2.557	12.53	-	-	-	-	-
SPO	2.052	10.06	-0.409	-2.64	-	-	0.409
OBB	1.755	8.69	-	-	0.417	2.50	-0.417
OBL	1.772	8.82	-	-	0.364	2.34	-0.364
ACC	2.216	10.78	-	-	-	-	-
RE	1.833	9.33	-0.359	-2.02	0.245	1.62	0.113
PUR	3.051	14.69	-	-	-	-	-
ST	2.036	10.08	-	-	-	-	-
CON U	3.450	16.63	-	-	-	-	-
CONI	3.870	18.25	-	-	-	-	-
LOC	3.488	16.55	0.430	2.66	-	-	-0.430
IAD	0.291	1.33	-	-	-	-	-
Log-Verosimilitud					-3057.83		
AIC/N					3.925		

Los coeficientes del modelo muestran las diferencias de cada grupo para cada atributo en particular. Para facilitar la visualización, los siguientes gráficos (Figura 23, Figura 24 y Figura 25) muestran las comparaciones de cada grupo con respecto al modelo estimado con la muestra completa (Tabla 71). Los gráficos están ordenados de mayor a menor frecuencia de uso del ferrocarril.

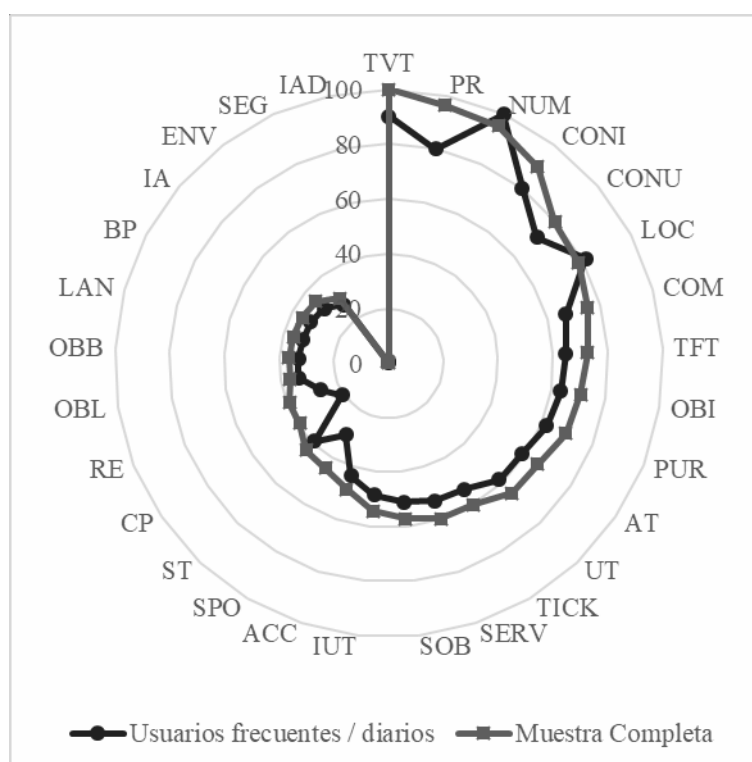


Figura 23 Comparativa entre la importancia de los atributos ferroviarios de la muestra completa y los usuarios frecuentes.

La principal diferencia entre los usuarios frecuentes y el resto (Figura 23) es la clara reducción de la importancia de los precios de los billetes (atributo PR). De la misma manera, el tiempo de viaje (TVT) reduce ligeramente su importancia, mientras que la

frecuencia de servicio (NUM) se convierte en la variable más importante. En cuanto a la accesibilidad y las conexiones con otros medios de transporte, ya sean urbanos (CONU) o interurbanos (CONI), la importancia se reduce ligeramente, mientras que la ubicación de las estaciones (LOC) se vuelve más importante. Al final de la lista, se observa una clara reducción de la importancia asociada a la seguridad policial (SPO), la disponibilidad de aparcamiento en las estaciones (CP) y la reserva de plazas (RE). A través de estas variaciones, podemos deducir que los usuarios regulares del ferrocarril son aquellos que utilizan el ferrocarril para los desplazamientos diarios. Por lo tanto, la importancia de los atributos se centra principalmente en tener una buena frecuencia de servicios y que las estaciones sean accesibles a sus orígenes y/o destinos.

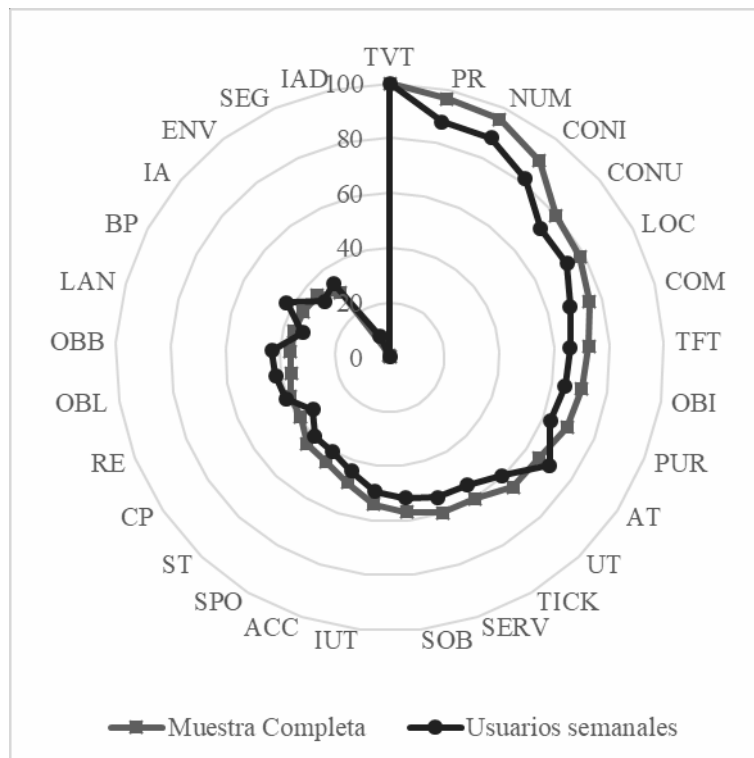


Figura 24 Comparativa entre la importancia de los atributos ferroviarios de la muestra completa y los usuarios semanales.

Los usuarios que utilizan el ferrocarril semanalmente (Figura 24) muestran una clara preferencia por el tiempo total de viaje (TVT), es decir, el tiempo de viaje en tren y el tiempo de acceso al mismo (AT). En relación con los servicios adicionales, se observa una mayor importancia de los servicios relacionados con la intermodalidad con la bicicleta (BP y OBB) y la posibilidad de llevar a bordo un gran volumen de equipaje (OBL). Podemos incluir en este grupo a las personas que realizan sus viajes diarios normalmente en bicicleta pero que la combinan con el ferrocarril por diferentes razones, como una mayor distancia de viaje. Por otro lado, este grupo también puede incluir a aquellas personas que realizan viajes de media y larga distancia y que optan por el ferrocarril si éste es competitivo con los otros modos de transporte en términos de tiempo de viaje y acceso.

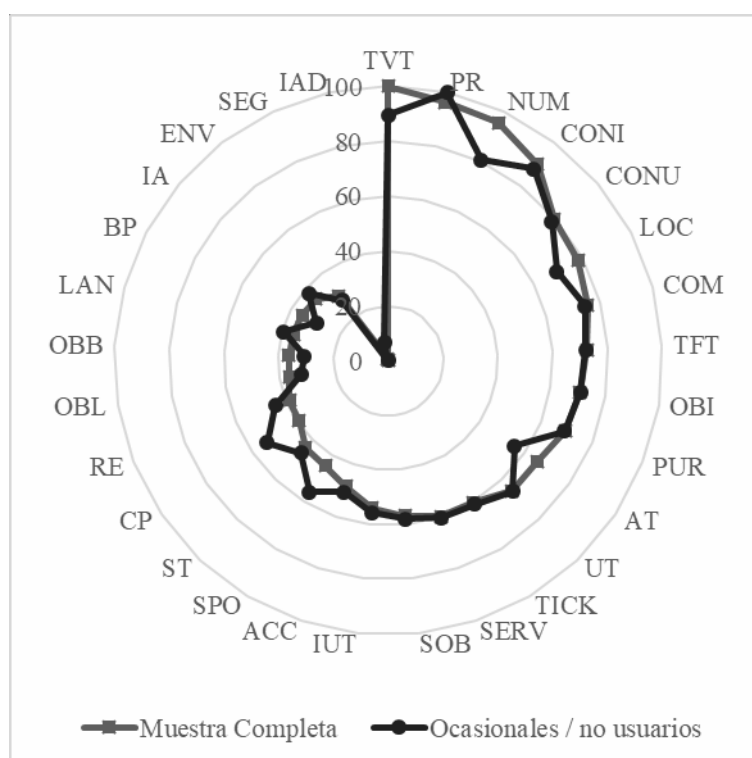


Figura 25 Comparativa entre la importancia de los atributos ferroviarios de la muestra completa y los usuarios ocasionales o no usuarios.

Por último, para las personas que nunca o casi nunca utilizan el ferrocarril (Figura 25), el precio (PR) tiene la mayor importancia. Además, la frecuencia (NUM), la ubicación de las estaciones (LOC) y el tiempo de acceso (AT) reducen considerablemente su importancia. En cambio, las conexiones urbanas (CONU) e interurbanas (CONI) con otros medios de transporte y la disponibilidad de aparcamiento en las estaciones (CP) aumentan su importancia. También se consideran más importantes los servicios adicionales centrados en los usuarios no habituales, como la disponibilidad de información en diferentes idiomas (LAN), la información a las estaciones de acceso (IA) o la información de las actividades en el lugar de destino (IAD). Dentro de este grupo, podemos incluir a los pasajeros que utilizan el ferrocarril si es económicamente más ventajoso que el resto de las alternativas. Además, estos usuarios aprecian los servicios adicionales que permiten acceder a la información necesaria del servicio.

“Cuando somos niños casi nunca pensamos en el futuro. Esa inocencia nos deja libres para disfrutar como pocos adultos pueden hacerlo. El día que empezamos a preocuparnos por el futuro es el día que dejamos atrás nuestra infancia.”

(Rothfuss, 2008)

CAPÍTULO 6: CONCLUSIONES FINALES Y LÍNEAS DE INVESTIGACIÓN FUTURAS

6.1 CONCLUSIONES ESPECÍFICAS

- En esta tesis se han desarrollado distintos métodos para analizar la calidad percibida por los usuarios de transporte público. Definir aquellos atributos que influyen en mayor o menor medida en la calidad global del servicio es de gran utilidad para las administraciones y operadores de los servicios. De esta forma, es posible concentrar las inversiones en aquellos aspectos más deficientes.
- Los modelos Ordered Probit han demostrado ser un método eficaz para estudiar la calidad percibida por los usuarios de transporte público. En especial, en lo referente a la posibilidad de captar la no-linealidad presente en las respuestas cualitativas. Quedando demostrado que no existe una misma distancia entre una valoración “Muy mal” a “Mal”, que entre “Bien” y “Muy bien”. El efecto de la no linealidad es capturado por los parámetros de límite μ .
- La inclusión de la importancia de los atributos, obtenida mediante ranking, en los modelos de datos ordenados no mejora de forma considerable el ajuste del modelo. El aspecto que sí es necesario tener en cuenta a la hora de modelizar es la heterogeneidad de los usuarios. Siendo la inclusión del mismo en el modelo un aspecto de mejora del ajuste importante. Considerando que, a la hora de realizar la primera encuesta de satisfacción presentada en esta tesis, la parte más larga de responder fue la parte precisamente referente al ranking de importancias de los atributos. De un punto de vista práctico, es posible decir que no es recomendable obtener los datos de importancia mediante un proceso de ranking. Por el contrario, es altamente recomendado obtener la información socioeconómica de los encuestados para utilizarlo posteriormente en el análisis de la calidad.
- En esta tesis se ha observado que, considerando la mitad de los datos disponibles, obtenidos en las encuestas de satisfacción, es posible realizar un análisis descriptivo de los niveles de satisfacción con una variación mínima. Obteniéndose diferencias menos al 3% en los promedios de los atributos y menores de 6% en el caso de las desviaciones estándar. Por lo tanto, se concluye que no es necesario obtener la satisfacción de todos los atributos por parte de todos los encuestados si el objetivo es estudiar únicamente la satisfacción de los usuarios. En consecuencia, las empresas operadoras pueden ahorrar recursos considerables al poder reducir la duración necesaria para responder cada encuesta, y, por ende, reducir el coste de toda la campaña de encuestado.
- Los resultados empíricos mostrados en esta tesis, han demostrado que es posible estimar modelos Ordered Probit partiendo de una base de datos parcial con una pérdida de información mínima. Para ello, el mejor método para completar los datos faltantes ha resultado ser la Imputación Múltiple. Obteniendo mediante este método resultados similares a los modelos estimados con la base de datos completa. Se ha realizado un test Vuong para modelos no anidados para confirmar que los modelos muestran un comportamiento similar. Más concretamente, el valor del test Vuong para el caso del modelo estimado con Imputación Múltiple ha sido de 0.39. Al tratarse este valor menor del límite considerado para un intervalo de confianza del 95% (1.96), se confirma la similitud entre modelos.
- Para establecer los atributos más relevantes que definen un sistema de transporte público se ha aplicado un método basado en técnicas cuantitativas y cualitativas.

Más concretamente, este método se ha empleado en ámbito ferroviario para conocer los atributos que mejor definen su servicio.

- Mediante el uso de grupos focales y entrevistas en profundidad se ha podido observar que la percepción del transporte ferroviario es diferente dependiendo del tipo de servicio.
- Tanto los expertos como los usuarios perciben que los servicios de alta velocidad disponen de frecuencia suficiente, servicios auxiliares adecuados e información apropiada en estaciones principales. Sin embargo, en estaciones secundarias o servicios de cercanías, la frecuencia y los servicios auxiliares son percibidos como deficientes.
- La automatización de los servicios ferroviarios, que puede aumentar en los próximos años, genera cierto nivel desconfianza en algunos usuarios. Generando una percepción de la atención al cliente demasiado impersonal.
- Las tarifas actuales en servicios ferroviarios se perciben como muy rígidas y no especialmente sensibles frente a distintos tipos de usuarios o viajes. Una mejora en la flexibilidad de las tarifas con un servicio puerta a puerta y en combinación con servicios low-cost podría atraer más usuarios.
- Mediante la aplicación del método Best-Worst, puede concluirse que la variable más importante en el transporte ferroviario es el tiempo de viaje total. En un segundo nivel se encuentran el precio y la frecuencia del servicio, que por el nivel de importancia que presentan deben acompañar a un tiempo de viaje adecuado. La accesibilidad al transporte ferroviario muestra una importancia considerable, centrándose principalmente en la intermodalidad del transporte ferroviario con el resto de modos de transporte, además de disponer de estaciones en buenas localizaciones dentro de las ciudades. Los servicios auxiliares muestran un nivel de importancia variado. Por ejemplo, la información a bordo y el confort de los asientos es un aspecto importante, mientras que la información sobre actividades en el destino, la posibilidad de comprar billetes en distintas clases o el poder viajar con mascotas no lo son.
- Para mejorar la calidad en el transporte ferroviario es necesario invertir en infraestructura y servicios además de actuar en la política tarifaria. Dejando de lado la inversión en servicios auxiliares que pueden ser apreciados por ciertos usuarios, pero no por la mayoría del público.
- El análisis realizado basado en distintos usuarios del ferrocarril ha mostrado variaciones en los niveles de los atributos. Los usuarios frecuentes (usuarios que utilizan el ferrocarril de forma diaria) no consideran el precio un atributo de mucha relevancia, mientras que la frecuencia y la localización de las estaciones gana más importancia. Los usuarios que utilizan el ferrocarril de forma esporádica consideran el tiempo de viaje como el aspecto clave. Para los usuarios menos frecuentes, o aquellos que no utilizan nunca el ferrocarril, el precio es el atributo más importante al igual que las conexiones de las estaciones y la posibilidad de aparcar en ellos. Los no usuarios del ferrocarril consideran los servicios auxiliares más importantes que el resto de usuarios.
- La forma de aumentar el número de pasajeros en el ferrocarril debería estar centrado en atraer usuarios poco frecuentes o no usuarios al mismo. Es altamente improbable que las medidas enfocadas a atraer este segmento de usuarios

desanimen al resto usuarios a seguir utilizando el ferrocarril. Por ello, la media más efectiva y rentable sería la de mejorar la intermodalidad del ferrocarril con los demás modos de transporte, tanto públicos como privados, y mejorar la información ofrecida a la población sobre los servicios disponibles.

- Se ha demostrado que el método de encuestado Best-Worst es adecuado para analizar tanto la satisfacción como la importancia de los usuarios.
- Los resultados de correlación han demostrado que el método tradicional (evaluación de la calidad con la posterior modelización mediante modelos de datos ordenados) y el método basado en encuestas de tipo Best-Worst son equivalentes a la hora de estudiar la satisfacción de los usuarios. El valor del coeficiente de correlación entre ambos métodos para la satisfacción de los distintos atributos ha resultado de 0.946.
- El modelo de regresión ha demostrado que es posible reproducir los resultados de la evaluación cualitativa mediante la escala Likert aplicando los resultados BW. El ajuste del modelo muestra un indicador de R^2 ajustado de 0.954. Este hecho tiene unas implicaciones importantes en cuanto a la eficiencia de encuestado. Puesto que las encuestas de tipo BW son más rápidas de realizar que las evaluaciones clásicas basadas en escalas de tipo Likert. Por lo tanto, este hecho sugiere que las largas encuestas basadas en evaluaciones pueden ser reemplazadas por preguntas de tipo BW.
- La importancia que cada variable tiene a la hora de explicar la calidad global del sistema es distinta dependiendo del tipo de modelización utilizado. Los resultados son consistentes con la literatura existente, donde se ha demostrado que existen diferencias entre la importancia explícita e implícita de los atributos. Más concretamente, mediante el método BW se han obtenido que los atributos de mayor importancia para los usuarios son el tiempo de viaje, la frecuencia del servicio y el precio, mientras que estos atributos no han resultado ser relevantes en los modelos de datos ordenados. El coeficiente de correlación entre ambos métodos es de 0.486, mostrando una cierta correlación entre ambos métodos. Los resultados obtenidos mediante el método BW han mostrado estar más en línea con los estudios existentes en la literatura, por esta razón, se considera que el método BW es adecuado para definir la importancia de los atributos. Mientras que el modelo Ordered (Logit o Probit) va más allá de definir el nivel de importancia. El modelo de regresión ha demostrado que los parámetros estimados en los modelos Ordered no solo están relacionados con la importancia de las variables, sino que también están afectados por su nivel de satisfacción.
- El análisis de Importancia - Rendimiento o *Importance – Performance Analysis* (IPA) muestra una clasificación que combina el nivel de satisfacción y la importancia de cada atributo. Es una herramienta muy útil para ayudar a los operadores de transporte público o autoridades a tomar decisiones de los aspectos a mejorar en su servicio. Dependiendo del método utilizado para establecer el nivel de satisfacción y la importancia los resultados pueden ser distintos. Los niveles de importancia establecidos mediante el método BW están más en línea con la literatura existente. Por lo tanto, el IPA basado en BW muestra unos resultados más apropiados sobre la prioridad de mejora de los atributos.

- Aunque los resultados de las encuestas BW han demostrado que existe la posibilidad de sustituir el método tradicional por el método BW, hay ciertos aspectos que hay que considerar. En primer lugar, para poder obtener el nivel de satisfacción de los atributos acotada en una escala, es necesario estimar el modelo de regresión que correlacione los resultados BW con los de la puntuación tradicional. Requiriendo para ello un estudio preliminar para estimar dicho modelo. Además, la satisfacción general del servicio debe ser evaluada mediante una escala tradicional independientemente del método utilizado para los atributos.

6.2 CONCLUSIONES GENERALES

Para estudiar la calidad en un sistema de transporte público es necesario aplicar métodos que combinen análisis cualitativo y cuantitativo. El análisis cualitativo sirve para establecer aquellas variables que mejor representan un sistema de transporte, los cuales son utilizados para desarrollar la encuesta de satisfacción. No realizar esta primera fase cualitativa puede afectar en que la longitud de la encuesta sea excesiva o que los atributos seleccionados no sean los adecuados para el transporte en cuestión. El análisis cuantitativo basado en modelización permite establecer las variables que más influyen en la satisfacción de los usuarios y ayudan a establecer políticas de mejora centrándose en aquellos aspectos más importantes.

Se ha observado que distintos métodos de modelado proporcionan resultados ligeramente distintos a la hora de establecer la importancia de los atributos. Sin embargo, el nivel de satisfacción de los usuarios con los atributos no varía dependiendo del sistema de análisis utilizado.

La obtención de datos realizada mediante encuestas de satisfacción supone la fase más costosa del análisis de la calidad percibida. El tiempo requerido para obtener los datos necesarios para la modelización puede reducirse mejorando el rendimiento de la fase de encuesta. Para ello se han estudiado dos formas. Por una parte, se ha considerado la posibilidad de estimar los modelos de elección discreta partiendo de una base de datos reducida. Concluyendo que es posible estimar los modelos disponiendo de la mitad de la información necesaria en condiciones normales. Para poder estimar los modelos es necesario completar la base de datos siguiendo un método de imputación, siendo la Imputación Múltiple la que mejores resultados a dado. La segunda forma de reducir la carga de las encuestas ha sido mediante el uso de métodos de encuesta alternativos. En esta tesis se ha utilizado el encuestado basado en preguntas de tipo Best-Worst, el cual ha permitido establecer la satisfacción y la importancia de los atributos de forma rápida planteando ejercicio de elección compuestos por 4 atributos. Los resultados referentes al nivel de satisfacción del método de valoración tradicional y el BW ha resultado ser muy similar. Sin embargo, la importancia derivada de los modelos ordered y la importancia basada en respuestas BW no ha resultado ser equivalente, si bien existe un alto nivel de correlación. En lo referente a la importancia, los resultados de la encuesta BW han mostrado estar más en consonancia con la bibliografía internacional.

La combinación de los resultados de satisfacción e importancia puede combinarse en un método IPA (Importance performance analysis), que permite diferenciar cuatro cuadrantes con atributos que son importantes y hay que mejorar, atributos que hay que mejorar pero no son importantes, atributos importantes que tienen un alto nivel de satisfacción y atributos no importantes con alto nivel de satisfacción. El análisis IPA resultante del análisis de calidad supone una herramienta muy útil para facilitar la toma de decisiones a los operadores de transporte público.

Resumiendo, los resultados de esta tesis han aportado distintas formas de aumentar el rendimiento de los procesos de encuestado en estudios satisfacción en sistemas de transporte público sin perjudicar o incluso mejorando los resultados de los mismos. Las técnicas desarrolladas en esta tesis son una aportación a la literatura y permitirán el desarrollo de nuevas investigaciones en el futuro.

6.3 LÍNEAS DE INVESTIGACIÓN FUTURA

Una tesis no sería una tesis si no acabase con más incógnitas que preguntas respondidas, y esta no es una excepción. Cada artículo presentado en este compendio añade más líneas de investigación que pueden desarrollarse en el futuro. Empezando por el primero, el método de ponderación utilizado para mejorar los modelos ha demostrado no tener un efecto demasiado elevado. Por ello, sería interesante utilizar algún otro valor para la ponderación basado en un método de encuesta distinto. Por ejemplo, los parámetros estimados mediante las encuestas Best-Worst podrían ser una buena alternativa.

Varios estudios presentados en esta tesis han demostrado que es importante considerar la heterogeneidad de la muestra para estudiar la calidad en el transporte público. El trabajo realizado hasta el momento se ha centrado en considerar las variaciones en los gustos de los usuarios considerando variables socioeconómicas y relacionadas con el viaje (sexo, edad, nivel de ingresos, motivo del viaje, etc.). Sin embargo, la satisfacción de los usuarios con un servicio de transporte no se limita solo a sus características personales. Los modelos muestran que los atributos de los servicios como tiempos de viaje o frecuencia son cuestiones importantes para los usuarios. Por esta razón queda estudiar la variabilidad en la satisfacción considerando la variación espacial y temporal de los servicios de transporte público. Dicho de otra forma, considerar en el estudio de la satisfacción los cambios que sufren los servicios de transporte público a lo largo del día, como reducciones de frecuencia en horas valle, o modificaciones en el recorrido de ciertas líneas a ciertas horas del día para dar acceso a ciertas zonas (por ejemplo, servicios especiales a universidades, colegios o eventos deportivos). Además, sería interesante considerar la diferencia del número de servicios en distintas zonas de una ciudad. Un barrio periférico puede disponer de una única línea que lo conecte con la ciudad con una frecuencia relativamente baja, mientras que una zona central de la ciudad dispone de un número de líneas muy superior con una frecuencia combinada alta. Todas estas variaciones en cuanto al servicio disponible pueden afectar a la satisfacción del usuario.

Se ha demostrado a lo largo de esta tesis que la realización de encuestas de satisfacción en el transporte público es lento y costoso. Los métodos de encuesta han ido variando a,

lo largo de los años llegando a realizarse en algunos casos de forma autónoma por el usuario. Sin embargo, la forma habitual de disponer de los datos, a día de hoy, sigue siendo mediante una encuesta más o menos extensa donde se evalúan todos los aspectos que interesan sobre el servicio. Dependiendo de la información requerida, las encuestas pueden requerir de varios minutos para ser completadas. Además, en la mayoría de los casos, las encuestas suelen tener un carácter general, esto es, no evalúan concretamente el servicio en el que se ha realizado la encuesta, sino que se evalúa el servicio en su conjunto o como mucho la línea encuestada. En esta tesis se han realizado varios esfuerzos en reducir la carga de preguntas necesarias en las encuestas, sin embargo, queda todavía trabajo por hacer de cara a optimizar la toma de datos. Los resultados relacionados con la posibilidad de realizar estudios completos partiendo de datos faltantes, o el uso de métodos basados en Best-Worst, han mostrado que no es necesario responder un cuestionario completo para estudiar la calidad de un sistema de transporte. Estos resultados pueden llevarse más adelante y desarrollar un método de encuesta más dinámico basado en métodos de micropulling.

Un sistema a desarrollar podría ser el siguiente. En primer lugar, se podrían instalar *beacons* bluetooth en los autobuses (u otros vehículos, dependiendo del modo de transporte analizado), asociando una ID de cada *beacon* al número de identificación del vehículo. Estas *beacons* permiten activar notificaciones en los teléfonos móviles de todos los usuarios que lleven instalada una App concreta al acercarse a los mismos. El usuario, al montarse (o desmontarse) al vehículo, recibiría una notificación pidiendo realizar la encuesta. En esta parte es donde los resultados de esta tesis entran en juego. La encuesta diseñada para esta app no podría ser una encuesta tradicional con muchas preguntas, sino que debería ser una encuesta reducida con 2 o 3 preguntas concretas. Por ejemplo, se podría preguntar el nivel de satisfacción de unos atributos concretos o mostrar ejercicios de elección de tipo BW. De forma complementaria, al estar basado, el sistema de encuesta en una aplicación móvil generalmente personal, sería posible preguntar de forma escalonada características socioeconómicas de los usuarios de forma que a través de varias encuestas sucesivas se pudiese completar el perfil de la persona encuestada. Por último, el análisis de los resultados se haría aplicando métodos similares a los explicados en esta tesis. Al realizarse las encuestas en el momento de realizar el viaje, los usuarios podrían evaluar el servicio utilizado de forma concreta, permitiendo un análisis más exhaustivo de los servicios de transporte público, observando aquellos momentos donde la calidad se ve reducida y por qué.

El método propuesto tiene sus limitaciones que son necesarias corregir si se quiere que este funcione. Los problemas son principalmente dos. El primero es que los usuarios deben tener instalada la aplicación móvil que permita activar las encuestas. Para ello, sería casi obligatorio que la app que activase las *beacons* fuera la misma aplicación de horarios utilizada por los usuarios, que incorporase la nueva funcionalidad mediante una actualización de software. Una nueva aplicación sin cuota de mercado que forzase al usuario a realizar encuestas de forma invasiva en sus dispositivos no resultaría eficaz. El segundo problema reside en que los usuarios quieran responder a las encuestas cada vez que utilicen un autobús. Para ello, sería muy útil establecer un sistema de incentivos donde las personas que respondiesen a las encuestas pudiesen disponer de saldo adicional en sus tarjetas de transporte para viajes futuros. Un sistema actualmente en el mercado

que permitiría entender el funcionamiento es la aplicación Google Opinion Rewards, que ofrece la posibilidad de responder a encuestas a cambio de saldo transferido a Google Wallet.

Hasta el momento solo nos hemos centrado en los aspectos a mejorar en el estudio de la satisfacción o la calidad percibida. Sin embargo, esa es solo una de las patas de la calidad en el transporte público. La calidad ofrecida, aquella controlada y gestionada por el operador no suele relacionarse con la calidad percibida por los usuarios. La calidad ofrecida se controla en mayor o menor medida en la práctica totalidad de cualquier sistema de transporte mínimamente estructurado y moderno. Mediante sistemas de geolocalización se controla a la flota para que cumplan con los mínimos de puntualidad, frecuencia u otros aspectos operativos. Una futura línea de investigación en la que ya se está trabajando se centra en estudiar la correlación existente entre la calidad ofrecida y la calidad percibida. De forma que sea posible estimar la satisfacción de los usuarios partiendo de los datos de control del operador, al estar estos generalmente disponibles de forma continua y en tiempo real.

Para entender las futuras tendencias en el estudio de la calidad en el transporte público, es necesario analizar las tendencias futuras en el transporte en general. Dejando de lado los posibles nuevos sistemas de transporte que se desarrollaran en el futuro con mayor o menor nivel de automatización, un aspecto en el que ya se han desarrollado varios pilotos es el concepto del MaaS (Mobility as a Service). El concepto del MaaS es en sí muy extenso y susceptible a distintas interpretaciones. Sin entrar en mucho detalle, lo que la mayoría de las corrientes indican es que el transporte del futuro estará muy interconectado entre sí llegando a viajes con un gran carácter intermodal, todo ello gestionado o coordinado por una entidad que engloba todos los modos disponibles. Esta es una limitación del sistema actual utilizado para estudiar la calidad. La norma UNE 13816 se centra en establecer un esquema de calidad para el transporte público, este esquema (ciclo de la calidad) explica aquellos aspectos que han de considerarse para tener un transporte público de calidad. Los modelos estimados basados en las encuestas sirven para establecer aquellos atributos que más afectan a la calidad y permiten establecer criterios de mejora útiles para tener un sistema de calidad. En ambos casos, todo el análisis se centra en un único sistema de transporte, considerando como mucho distintas líneas de un mismo sistema. Actualmente, se están desarrollando distintos estudios que analizan viajes intermodales, sin embargo, esto se centran en combinaciones concretas sin considerar un esquema más general como puede acabar desarrollándose en un concepto MaaS. Un esquema completo de calidad supondría disponer de ciclos de calidad individuales para cada uno de los sistemas incluidos en el MaaS, a su vez, la entidad que gestionase el sistema tendría su propio ciclo de calidad. A su vez, la modelización conjunta del sistema sería más compleja al ser específica de cada viaje, puesto que cada usuario encadenaría distintos modos de transporte según más le convenga. Cada modo de transporte tendría asociados unos atributos representativos que pueden o no estar interrelacionados, además, la plataforma de gestión de los transportes sería también evaluada y la intermodalidad de los pares de modos concretos utilizados. De todas formas, aunque es posible desarrollar un marco teórico de modelización, sería aconsejable esperar hasta que los servicios MaaS estén completamente implementados para poder empezar a evaluar la satisfacción de los usuarios.

“When we are children we seldom think of the future. This innocence leaves us free to enjoy ourselves as few adults can. The day we fret about the future is the day we leave our childhood behind.”

(Rothfuss, 2008)

CAPÍTULO 7: CONCLUSIONS AND FUTURE LINES OF RESEARCH

7.1 SPECIFIC CONCLUSIONS

- In this thesis, different methods to analyse the quality perceived by public transport users have been developed. Defining those attributes that influence to a greater or lesser extent the overall quality of the service is very useful for administrations and service operators. This way, it is possible to concentrate investments on those aspects that are most deficient.
- Ordered Probit models have proven to be an effective method for studying the perceived quality in public transport systems. In particular, with regard to the possibility of capturing the non-linearity present in qualitative responses. It has been demonstrated that there is not the same distance between a "Very bad" and "Bad" rating as between "Good" and "Very good". Is indeed more difficult to improve a score when the rating is higher. The effect of non-linearity is captured by the threshold parameters μ .
- The inclusion of the importance of the attributes, obtained through ranking, in the ordered data models does not significantly improve model's fit. The aspect that does need to be taken into account when modelling is the heterogeneity of the sample. The inclusion of user heterogeneity in the model is an important aspect of fit improvement. Considering that, at the time of carrying out the first satisfaction survey presented in this thesis, the most time consuming part of the survey was the attribute ranking task. From a practical point of view, it is possible to say that it is not advisable to obtain the importance data through a ranking process. On the contrary, it is highly recommended to obtain the socioeconomic information from the respondents to be used later in the quality analysis.
- In this thesis it has been observed that, using only half of the available data obtained in the satisfaction surveys, it is possible to carry out a descriptive analysis study of the satisfaction levels with a minimum variation on results. Differences of less than 3% are obtained in the averages of the attributes and less than 6% in the case of the standard deviations. Therefore, it is concluded that it is not necessary to obtain the satisfaction rating of all attributes from all respondents if the objective is to study only satisfaction levels. Asking only half of the rating questions to each respondent have been proven to be enough. Consequently, operating companies can save considerable resources by being able to reduce the length of time required to respond each survey, thereby reducing the cost of the entire survey campaign.
- The empirical results shown in this thesis have demonstrated that it is possible to estimate Ordered Probit models from a partial database with a minimum loss of information. For this purpose, the best method to complete the missing data has turned out to be Multiple Imputation. Obtaining by this method similar results to the models estimated with the complete database. A Vuong test for non-nested models has been performed to confirm that the models show similar behaviour. More specifically, the Vuong test value for the case of the model estimated with Multiple Imputation is 0.39. As this value is lower than the limit considered for a 95% confidence interval (1.96), the similarity between models is confirmed.
- A method based on quantitative and qualitative techniques has been applied to establish the most relevant attributes to define a transport system. More

specifically, this method has been used in the railway field to find out the attributes that best define its service.

- Through the use of focus groups and in-depth interviews it has been possible to observe that the perception of rail transport is different depending on the type of service (urban, intercity or long distance / high speed).
- Both experts and users perceive that high-speed services have sufficient frequency, adequate auxiliary services and appropriate information at main stations. However, in secondary stations or intercity services the frequency and auxiliary services are perceived as deficient.
- The automation of railway services, which may increase in the coming years, generates a certain level of mistrust in some users. It generates a perception of too impersonal customer service.
- The current fares for rail services are perceived as very rigid and not particularly sensitive to different types of users or trips. An improvement in fare flexibility with a door-to-door service and in combination with low-cost services could attract more users.
- By applying the Best-Worst method, it can be concluded that the most important variable in rail transport is total travel time. On a second level are the price and frequency of the service, which, because of their importance, must be accompanying the travel time at the same level. Accessibility to rail transport shows considerable importance, focusing mainly on the intermodality of rail transport with other modes of transport, in addition to having stations in good locations within cities. Ancillary services show a varied level of importance. For example, information on board and seat comfort is an important aspect, while information on activities at the destination, the possibility of buying class segregated tickets or being able to travel with pets are not.
- Improving quality in rail transport requires investment in infrastructure and services as well as action on pricing policy. Leaving aside investment in ancillary services that may be appreciated by certain users, but not by the majority of the public.
- Analysis based on different railway users has shown variations in the attribute importance levels. Frequent users (those who use the railway on a daily basis) do not consider price as a very relevant attribute, while frequency and location of stations gain more importance. Users who use the train sporadically consider travel time as the key aspect. For less frequent users, or those who never use the train, price is the most important attribute as well as station connections and the possibility of parking at stations. Non-rail users consider ancillary services more important than other users.
- To increase the number of passengers on railways, investments should focus on attracting infrequent users or non-users. It is highly unlikely that measures aimed at attracting this segment of users will discourage other users from continuing to use the railways. Therefore, the cost-effective means would be to improve intermodality between rail and other modes of transport, both public and private, and to improve the information given to people about the available services.
- The Best-Worst survey method has been shown to be suitable for analysing both user satisfaction and importance.

- Correlation results have shown that the traditional method (quality assessment with subsequent modelling using ordered models) and the Best-Worst survey-based method are equivalent when studying user satisfaction. The value of the correlation coefficient between the two methods for the satisfaction of the different attributes is 0.946.
- The regression model has shown that it is possible to reproduce the results of the qualitative questionnaires using the Likert scale by applying the BW results. The model's fit shows an adjusted R^2 indicator of 0.954. This fact has important implications for the efficiency of respondents. Since BW-type surveys are faster to conduct than classical surveys based on Likert scales. Therefore, this fact suggests that long surveys based on scores can be replaced by BW-type questions.
- The importance that each variable has in explaining the overall quality of the system is different depending on the model used. The results are consistent with the existing literature, where it has been shown that there are differences between the explicit and implicit importance of the attributes. More specifically, using the BW method it has been obtained that the attributes of greatest importance for users are travel time, service frequency and price, while these attributes have not been found to be relevant in the ordered models. The correlation coefficient between the two methods is 0.486, showing some level of correlation between the two methods. The results obtained by the BW method have been shown to be more in line with existing studies in the literature, for this reason, the BW method is considered adequate to define the importance of the attributes, while the parameters estimated with the Ordered models (Logit or Probit) go beyond defining just the level of importance. The regression model has shown that the parameters estimated in Ordered models are not only related to the importance of the variables but are also affected by their level of satisfaction.
- The Importance - Performance Analysis (IPA) shows a classification that combines the level of satisfaction and the importance of each attribute. It is a very useful tool to help public transport operators or authorities to make decisions about the aspects to improve in their services. Depending on the method used to establish the level of satisfaction and importance the results can be different. The levels of importance established by the BW method are more in line with existing literature. Therefore, the IPA based on BW shows more appropriate results on the priority of improvement of the attributes.
- Although the results of the BW surveys have shown that there is a possibility of replacing the traditional method with the BW scaling, there are certain aspects to be considered. Firstly, in order to obtain the level of satisfaction of the attributes in a scale, it is necessary to estimate the regression model that correlates the BW results with those of the traditional score. This requires a preliminary study to estimate this model. In addition, overall service satisfaction should be assessed using a traditional Likert scale regardless of the method used for the attributes.

7.2 GENERAL CONCLUSIONS

To study quality in a public transport system it is necessary to apply methods that combine qualitative and quantitative analysis. The qualitative analysis serves to establish those variables that best represent a transport system, which are used to develop the satisfaction survey. Not carrying out this first qualitative phase can affect the length of the survey or the suitability of the selected attributes for the transport in question. Quantitative analysis based on modelling allows the establishment of the variables that most influence user satisfaction and help to establish improvement policies focusing on the most important aspects.

It has been observed that different modelling methods give slightly different results when establishing the importance of attributes. However, the level of user satisfaction with the attributes does not vary depending on the analysis system used.

Obtaining data through satisfaction surveys is the most expensive phase of the perceived quality analysis. The time required to obtain the data needed for modelling can be reduced by improving the performance of the survey phase. Two ways have been studied to do this. On the one hand, the possibility of estimating discrete choice models from a reduced database has been considered. The conclusion is that it is possible to estimate the models by having only half of the necessary information available under normal conditions. To be able to estimate the models it is necessary to complete the database following an imputation method, being the Multiple Imputation the one that has given better results. The second way to reduce the burden of the surveys has been by using alternative survey methods. In this thesis, we have used a survey based on Best-Worst scaling questions. This has allowed us to establish the satisfaction and importance levels of each attribute faster by proposing a choice exercise composed of 4 attributes. The results regarding the level of satisfaction of the traditional rating method and the BW scaling have been very similar. However, the importance derived from the ordered models and the importance based on BW responses has not proved to be equivalent, although there is a high level of correlation between both results. In terms of significance, the results of the BW survey have been shown to be more in line with the international literature.

The combination of satisfaction and importance results can be combined in an IPA (Importance performance analysis) chart. Four quadrants are defined: (i) attributes that are important and need to be improved, (ii) attributes that need to be improved but are not important, (iii) important attributes that have a high level of satisfaction and (iv) unimportant attributes with a high level of satisfaction. The attributes located in the first quadrant are those with high improvement priority. The IPA chart resulting from the quality analysis is a very useful tool to facilitate decision making to public transport operators.

In summary, the results of this thesis have provided different ways to increase the performance of the survey processes in satisfaction studies in public transport systems without damaging or even improving the results. The techniques developed in this thesis are a contribution to the literature and will allow the development of new research in the future.

7.3 FUTURE LINES OF RESEARCH

A thesis would not be a thesis if it did not end with more open than closed questions, and this one is not an exception. Each article presented in this compendium adds more lines of research that can be developed in the future. Starting with the first one, the weighting method used to improve the models has proven not to be very effective. It would therefore be interesting to use some other value for weighting based on a different respondent method. For example, parameters estimated using the Best-Worst survey could be a good alternative.

Several studies presented in this thesis have shown that it is important to consider the heterogeneity of the sample in order to study quality in public transport. The work carried out so far has focused on considering variations in user tastes by considering socio-economic and travel-related variables (gender, age, income level, reason for the trip, etc.). However, user satisfaction towards a transport service is not limited only to personal characteristics. The models show that service attributes such as travel times or frequency are important issues for users. For this reason, it remains to study the variability in satisfaction by considering the spatial and time variation of public transport services. In other words, the study of satisfaction should consider the changes that public transport services undergo throughout the day, such as reductions in frequencies during off-peak hours, modifications in the route of certain lines at certain times of the day to provide access to certain areas (for example, special services to universities or schools). In addition, it would be interesting to consider the difference in the number of services available in different areas of a city. An outlying neighbourhood may have a single line connecting it to the city centre with a relatively low frequency, while a central area of the city has a much higher number of lines with a high combined frequency. All these variations in the service variables can affect user satisfaction.

It has been demonstrated throughout this thesis that carrying out satisfaction surveys in public transport is slow and costly. The methods used to carry out the surveys have varied over the years, and in some cases they have been carried out independently by the user. However, the usual way of having the data, nowadays, is still by carrying out a more or less lengthy survey where all the interesting aspects of the service are evaluated. Depending on the information required, surveys may take several minutes to complete. In addition, in most cases, surveys tend to be general in nature, that is, they do not specifically evaluate the service on which the survey has been conducted, but rather evaluate the service as a whole, or at most, focused on the surveyed line. In this thesis, several efforts have been made to reduce the time required to complete surveys. However, there is still work to be done in order to optimise the data collection process. The results related to the possibility of carrying out complete studies from missing data, or the use of Best-Worst based methods, have shown that it is not necessary to answer a complete questionnaire to study the quality of a transport system. These results can be taken further and a more dynamic survey method based on micropulling methods can be developed.

One dynamic system to be developed could be the following. Firstly, bluetooth beacons could be installed in buses (or other vehicles, depending on the mode of transport analysed), associating the ID of each beacon to the specific identification number of the

vehicle. These beacons allow to activate notifications in the mobile phones of all users with a specific App installed when approaching them. The user, when getting on (or of) the vehicle, would receive a notification asking them to carry out a survey. This is where the results of this thesis come into play. The survey designed for this app could not be a traditional survey with a lot of questions, but should be a reduced survey with 2 or 3 specific questions. For example, the level of satisfaction of a few attributes or a BW-type choice exercise. As a complement, since the survey system is based on a mobile application, generally personal and non-transferable, it would be possible to ask users' socioeconomic characteristics in a staggered way so that the profile of the person could be completed through several successive surveys. Finally, the analysis of the results would be done by applying similar methods to those explained in this thesis. By carrying out the surveys when the trips are made, users could evaluate the service in a specific way, allowing a more exhaustive analysis of public transport services, observing those moments when quality is reduced and why.

The proposed method has its limitations that need to be corrected if it is to work. The problems are mainly two. The first is that users must have the app installed to activate the surveys. To do this, it would be almost mandatory that the app that activates the beacons is the same timetable/trip planner app used by the users, which has to incorporate the new functionality by a software update. A brand new application without market share that would force the user to carry out surveys in an invasive way on their devices would not be effective. The second problem is that users may not want to respond to surveys every time they use a bus. To do this, it would be very useful to establish an incentive system where respondents could get additional balance on their transport cards for future trips. One system currently on the market that could help understand how this works is the Google Opinion Rewards application, which offers the possibility of responding surveys in exchange for money that is transferred to Google Wallet.

So far we have only focused on the aspects to be improved in the study of satisfaction or perceived quality. However, that is only one of the legs of quality in public transport. The offered or delivered quality, that which is controlled and managed by the operator, is not usually connected to the quality perceived by users. The offered quality is controlled to a greater or lesser extent in almost every minimally structured and modern transport system. Using GPS systems, the fleet is controlled so that they comply with the minimums of punctuality, frequency or other operational aspects. A future line of research on which work is already underway focuses on studying the correlation between the offered quality and the perceived. This will make it possible to estimate user satisfaction based on the operator's control data, as these is generally available on a continuous basis and in real time.

In order to understand future trends in the study of quality in public transport, it is necessary to analyse future trends in transport in general. Leaving aside the possible new transport systems that will be developed in the future with a greater or lesser level of automation, one aspect in which several pilots have already been developed is the concept of MaaS (Mobility as a Service). The MaaS concept itself is very extensive and open to different interpretations. Without going into too much detail, what most trends indicate is that the transport of the future will be highly interconnected, leading to journeys with a strong intermodal character, all managed or coordinated by an entity that encompasses all

the available modes. This is a limitation of the current system of studying quality. The EN 13816 standard focuses on establishing a quality scheme for public transport systems. This scheme (quality cycle) explains those aspects that must be considered in order to have high quality standards in public transport. The estimated models based on the surveys serve to establish those attributes that are most important and help to define an improvement criteria in order to have a high quality system. In both cases, the entire analysis focuses on a single transport system, considering at most different lines of the same system. Currently, different studies are being developed that analyse intermodal trips, however, this are focused on specific combinations without considering a more general scheme as it could end up developing into a MaaS concept. A complete quality scheme would mean having individual quality cycles for each of the systems included in the MaaS. In turn, the entity managing the system would have its own quality cycle. Furthermore, the joint modelling of the system would be more complex, as it would be specific to each journey, since each user would chain different modes of transport according to their own convenience. Each mode of transport would have associated representative attributes that may or may not be interrelated with other modes. The transport management platform would also be assessed and same for the intermodality of the specific mode pairs used. However, although it is possible to develop a theoretical modelling framework, it will be advisable to wait until the MaaS services are fully implemented before being able to start assessing user satisfaction properly.

REFERENCIAS

- Abenoza, R.F., Cats, O., Susilo, Y.O., 2017. Travel satisfaction with public transport: Determinants, user classes, regional disparities and their evolution. *Transp. Res. Part A Policy Pract.* 95, 64–84. <https://doi.org/10.1016/j.tra.2016.11.011>
- Ahern, A.A., Tapley, N., 2008. The use of stated preference techniques to model modal choices on interurban trips in Ireland. *Transp. Res. Part A Policy Pract.* 42, 15–27. <https://doi.org/https://doi.org/10.1016/j.tra.2007.06.005>
- Akaike, H., 1973. Information Theory and An Extension of the Maximum Likelihood Principle. In *Second International Symposium on Information Theory*, ed. B. N. Petrov and F. Csaki 22, 267–281.
- Albalade, D., Bel, G., Fageda, X., 2015. Competition and cooperation between high-speed rail and air transportation services in Europe. *J. Transp. Geogr.* 42, 166–174. <https://doi.org/https://doi.org/10.1016/j.jtrangeo.2014.07.003>
- Alegria, M., Takeuchi, D., Canino, G., Duan, N., Shrout, P., Meng, X.-L., Vega, W., Zane, N., Vila, D., Woo, M., Vera, M., Guarnaccia, P., Aguilar-Gaxiola, S., Sue, S., Escobar, J., Lin, K.-M., Gong, F., 2004. Considering Context, Place, and Culture: The National Latino and Asian American Study. *Int J Methods Psychiatr Res. Int J Methods Psychiatr Res* 13, 208–220.
- Allen, J., Eboli, L., Forciniti, C., Mazzulla, G., Ortúzar, J. de D., 2019. The role of critical incidents and involvement in transit satisfaction and loyalty. *Transp. Policy* 75, 57–69. <https://doi.org/10.1016/j.tranpol.2019.01.005>
- Allen, J., Eboli, L., Mazzulla, G., Ortúzar, J. de D., 2018a. Effect of critical incidents on public transport satisfaction and loyalty: an Ordinal Probit SEM-MIMIC approach. *Transportation (Amst)*. 1–37. <https://doi.org/10.1007/s11116-018-9921-4>
- Allen, J., Muñoz, J.C., Ortúzar, J. de D., 2018b. Modelling service-specific and global transit satisfaction under travel and user heterogeneity. *Transp. Res. Part A Policy Pract.* 113, 509–528. <https://doi.org/10.1016/j.tra.2018.05.009>
- Allison, P.D., 2000. Multiple imputation for missing data: A cautionary tale. *Sociol. Methods Res.* 28, 301–309. <https://doi.org/10.1177/0049124100028003003>
- Alonso, B., Barreda, R., Dell’Olio, L., Ibeas, A., 2018. Modelling user perception of taxi service quality. *Transp. Policy* 63, 157–164. <https://doi.org/10.1016/j.tranpol.2017.12.011>
- Andreassen, T.W., 1995. (Dis)satisfaction with public services: the case of public transportation. *J. Serv. Mark.* 9, 30–41. <https://doi.org/doi:10.1108/08876049510100290>
- Aydin, N., Celik, E., Gumus, A.T., 2015. A hierarchical customer satisfaction framework for evaluating rail transit systems of Istanbul. *Transp. Res. Part A Policy Pract.* 77, 61–81. <https://doi.org/10.1016/j.tra.2015.03.029>
- Baltes, M.R., 2003. The Importance Customers Place on Specific Service Elements of Bus Rapid Transit. *Transit* 6, 1–19. <https://doi.org/10.5038/2375-0901.6.4.1>

- Beck, M.J., Rose, J.M., 2016. The best of times and the worst of times: A new best-worst measure of attitudes toward public transport experiences. *Transp. Res. Part A Policy Pract.* 86, 108–123. <https://doi.org/10.1016/j.tra.2016.02.002>
- Beck, M.J., Rose, J.M., Greaves, S.P., 2017. I can't believe your attitude: a joint estimation of best worst attitudes and electric vehicle choice. *Transportation (Amst.)* 44, 753–772. <https://doi.org/10.1007/s11116-016-9675-9>
- Berry, L.L., Parasuraman, A., Zeithaml, V.A., 1990. Delivering Quality Service: Balancing Customer Perceptions and Expectations. *Business*. <https://doi.org/10.1177/0001699303046002008>
- Boes, S., Winkelmann, R., 2006. Ordered Response Models, in: *Analysis of Microdata*. Springer, pp. 171–205. https://doi.org/10.1007/3-540-29607-7_6
- Bolton, R.N., Drew, J.H., 1991. A multi-stage model of consumers' assessments of service quality and value. *J. Consum. Res.* 17, 375. <https://doi.org/10.1086/208564>
- Bordagaray, M., Dell'Olio, L., Ibeas, A., Cecín, P., 2014. Modelling user perception of bus transit quality considering user and service heterogeneity. *Transp. A Transp. Sci.* 10, 705–721. <https://doi.org/10.1080/23249935.2013.823579>
- Bošković, B., Bugarinović, M., 2015. Why and how to manage the process of liberalization of a regional railway market: South-Eastern European case study. *Transp. Policy* 41, 50–59. <https://doi.org/https://doi.org/10.1016/j.tranpol.2015.03.009>
- Box, G.E.P., Luceño, A., Paniagua-Quñones, M. del C., 2009. Statistical control by monitoring and adjustment. John Wiley & Sons.
- Brand, J., 1999. Development , implementation and evaluation of multiple imputation strategies for the statistical analysis of incomplete data sets.
- Brand, J.P.L., Van Buuren, S., Groothuis-Oudshoorn, K., Gelsema, E.S., 2003. A toolkit in SAS for the evaluation of multiple imputation methods. *Stat. Neerl.* 57, 36–45. <https://doi.org/10.1111/1467-9574.00219>
- Brons, M., Givoni, M., Rietveld, P., 2009. Access to railway stations and its potential in increasing rail use. *Transp. Res. Part A Policy Pract.* 43, 136–149. <https://doi.org/https://doi.org/10.1016/j.tra.2008.08.002>
- Bruseberg, A., McDonagh-Philp, D., 2002. Focus groups to support the industrial/product designer: a review based on current literature and designers' feedback. *Appl. Ergon.* 33, 27–38. [https://doi.org/10.1016/S0003-6870\(01\)00053-9](https://doi.org/10.1016/S0003-6870(01)00053-9)
- Burlando, C., Ivaldi, E., Musso, E., 2016. An indicator for measuring the perceived quality of local public transport: Relationship with use and satisfaction with the ticket price. *Int. J. Transp. Econ.* 43. <https://doi.org/10.19272/201606704003>
- Burton, A., Billingham, L.J., Bryan, S., 2007. Cost-effectiveness in clinical trials: using multiple imputation to deal with incomplete cost data. *Clin. Trials J. Soc. Clin. Trials* 4, 154–161. <https://doi.org/10.1177/1740774507076914>
- Cao, J., Cao, X., 2017. Comparing importance-performance analysis and three-factor theory in assessing rider satisfaction with transit. *J. Transp. Land Use* 10. <https://doi.org/10.5198/jtlu.2017.907>

- Cascetta, E., Cartenì, A., 2014. A Quality-Based Approach to Public Transportation Planning: Theory and a Case Study. *Int. J. Sustain. Transp.* 8, 84–106. <https://doi.org/10.1080/15568318.2012.758532>
- Casella, G., George, E.I., Casella, G., George, E.I., 2016. Explaining the Gibbs Sampler Stable URL : <http://www.jstor.org/stable/2685208> Linked references are available on JSTOR for this article : Explaining the Gibbs Sampler 3, 167–174.
- Cats, O., Abenoza, R.F., Liu, C., Susilo, Y.O., 2015. Evolution of satisfaction with public transport and its determinants in Sweden identifying priority areas, *Transportation Research Record*. <https://doi.org/10.3141/2538-10>
- Çelik, A.K., Senger, Ö., 2016. A COMPARISON OF ALTERNATIVE ORDERED RESPONSE MODELS FOR ANALYSING USER SATISFACTION WITH TRANSPORT SERVICES IN TURKEY. *Transp. Telecommun.* 17, 40–54. <https://doi.org/10.1515/ttj-2016-0005>
- Central Intelligence Agency, 2017. *The World Factbook 2017*.
- Chiou, J.-M., Zhang, Y.-C., Chen, W.-H., Chang, C.-W., 2014. A functional data approach to missing value imputation and outlier detection for traffic flow data. *Transp. B Transp. Dyn.* 2, 106–129. <https://doi.org/10.1080/21680566.2014.892847>
- Cook, W.D., Kress, M., 1988. Deriving weights from pairwise comparison ratio matrices: An axiomatic approach. *Eur. J. Oper. Res.* 37, 355–362. [https://doi.org/10.1016/0377-2217\(88\)90198-1](https://doi.org/10.1016/0377-2217(88)90198-1)
- D'Alfonso, T., Jiang, C., Bracaglia, V., 2015. Would competition between air transport and high-speed rail benefit environment and social welfare? *Transp. Res. Part B Methodol.* 74, 118–137. <https://doi.org/https://doi.org/10.1016/j.trb.2015.01.007>
- Das, T., Apu, N., Hoque, M.S.S., Hadiuzzaman, M., Xu, W., 2017. Parameters Affecting the Overall Performance of Bus Network System at Different Operating Conditions: A Structural Equation Approach, in: *Transportation Research Procedia*. Elsevier B.V., pp. 5063–5075. <https://doi.org/10.1016/j.trpro.2017.05.206>
- de Oña, J., de Oña, R., 2015. Quality of Service in Public Transport Based on Customer Satisfaction Surveys: A Review and Assessment of Methodological Approaches. *Transp. Sci.* 49, 605–622. <https://doi.org/10.1287/trsc.2014.0544>
- de Oña, J., de Oña, R., Calvo, F.J., 2012. A classification tree approach to identify key factors of transit service quality. *Expert Syst. Appl.* 39, 11164–11171. <https://doi.org/10.1016/j.eswa.2012.03.037>
- de Oña, J., de Oña, R., Diez-mesa, F., Eboli, L., Mazzulla, G., 2016a. A Composite Index for Evaluating Transit Service Quality across Different User Profiles. *J. Public Transp.* 19, 128–153. <https://doi.org/10.5038/2375-0901.19.2.8>
- de Oña, J., de Oña, R., Eboli, L., Forciniti, C., Mazzulla, G., 2016b. Transit passengers' behavioural intentions: the influence of service quality and customer satisfaction. *Transp. A Transp. Sci.* 12, 385–412. <https://doi.org/10.1080/23249935.2016.1146365>
- de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2016c. Index numbers for monitoring transit service quality. *Transp. Res. Part A Policy Pract.* 84, 18–30. <https://doi.org/10.1016/j.tra.2015.05.018>

- de Oña, J., de Oña, R., Eboli, L., Mazzulla, G., 2013. Perceived service quality in bus transit service: A structural equation approach. *Transp. Policy* 29, 219–226. <https://doi.org/10.1016/j.tranpol.2013.07.001>
- de Oña, J., de Oña, R., Garrido, C., 2017. Extraction of attribute importance from satisfaction surveys with data mining techniques: a comparison between neural networks and decision trees. *Transp. Lett.* 9, 39–48. <https://doi.org/10.1080/19427867.2015.1136917>
- de Oña, J., de Oña, R., López, G., 2016d. Transit service quality analysis using cluster analysis and decision trees: a step forward to personalized marketing in public transportation. *Transportation (Amst)*. 43, 725–747. <https://doi.org/10.1007/s11116-015-9615-0>
- de Oña, R., Eboli, L., Mazzulla, G., 2014. Key factors affecting rail service quality in the Northern Italy: a decision tree approach. *Transport* 29, 75–83. <https://doi.org/10.3846/16484142.2014.898216>
- de Oña, R., Machado, J.L., de Oña, J., 2015. Perceived Service Quality, Customer Satisfaction, and Behavioral Intentions. *Transp. Res. Rec. J. Transp. Res. Board* 2538, 76–85. <https://doi.org/10.3141/2538-09>
- del Castillo, J.M., Benitez, F.G., 2012. Determining a public transport satisfaction index from user surveys. *Transportmetrica* 9, 1–29. <https://doi.org/10.1080/18128602.2011.654139>
- dell'Olio, L., Ibeas, A., Barreda, R., Sañudo, R., Sañudo, R., 2013. Passenger behavior in trains during emergency situations. *J. Safety Res.* 46, 157–166. <https://doi.org/10.1016/j.jsr.2013.05.005>
- dell'Olio, L., Ibeas, A., Cecin, P., 2011. The quality of service desired by public transport users. *Transp. Policy* 18, 217–227. <https://doi.org/10.1016/j.tranpol.2010.08.005>
- dell'Olio, L., Ibeas, A., Cecin, P., 2010. Modelling user perception of bus transit quality. *Transp. Policy* 17, 388–397. <https://doi.org/10.1016/j.tranpol.2010.04.006>
- dell'Olio, L., Ibeas, A., de Oña, J., de Oña, R., 2018. Public Transportation Quality of Service. pp. 7–32.
- dell'Olio, L., Ibeas, A., de Oña, J., de Oña, R., 2017. Public Transportation Quality of Service: Factors, Models, and Applications, Elsevier. Elsevier, Cambridge, MA. <https://doi.org/10.1080/01441647.2018.1531083>
- Diana, M., 2012. Measuring the satisfaction of multimodal travelers for local transit services in different urban contexts. *Transp. Res. Part A Policy Pract.* 46, 1–11. <https://doi.org/10.1016/j.tra.2011.09.018>
- Domencich, T.A., McFadden, D., 1975. Urban Travel Demand: A Behavioral Analysis. North-Holland.
- Donders, A.R.T., van der Heijden, G.J.M.G., Stijnen, T., Moons, K.G.M., 2006. Review: A gentle introduction to imputation of missing values. *J. Clin. Epidemiol.* 59, 1087–1091. <https://doi.org/10.1016/j.jclinepi.2006.01.014>
- Eboli, L., Forciniti, C., Mazzulla, G., 2018. Spatial variation of the perceived transit service quality at rail stations. *Transp. Res. Part A Policy Pract.* 114, 67–83.

- <https://doi.org/10.1016/j.tra.2018.01.032>
- Eboli, L., Mazzulla, G., 2009. A new customer satisfaction index for evaluating transit service quality. *J. Public Transp.* 12, 21–37. <https://doi.org/http://dx.doi.org/10.5038/2375-0901.12.3.2>
- Eboli, L., Mazzulla, G., 2015. Relationships between rail passengers satisfaction and service quality: a framework for identifying key service factors. *Public Transp.* 7, 185–201. <https://doi.org/10.1007/s12469-014-0096-x>
- Eboli, L., Mazzulla, G., 2012. Performance indicators for an objective measure of public transport service quality. *Eur. Transp. \ Trasp. Eur.* 51, 1825–3997.
- Eboli, L., Mazzulla, G., 2011. A methodology for evaluating transit service quality based on subjective and objective measures from the passenger's point of view. *Transp. Policy* 18, 172–181. <https://doi.org/10.1016/j.tranpol.2010.07.007>
- Eboli, L., Mazzulla, G., 2010. How to capture the passengers' point of view on a transit service through rating and choice options. *Transp. Rev.* 30, 435–450. <https://doi.org/10.1080/01441640903068441>
- Eboli, L., Mazzulla, G., 2007. Service Quality Attributes Affecting Customer Satisfaction for Bus Transit. *J. Public Transp.* 10, 21–34. <https://doi.org/10.5038/2375-0901.10.3.2>
- EC, 1999. Quattro Final Report, Transport. ed. Paris.
- Echaniz, E., dell'Olio, L., Ibeas, Á., 2018. Modelling perceived quality for urban public transport systems using weighted variables and random parameters. *Transp. Policy* 67, 31–39. <https://doi.org/10.1016/j.tranpol.2017.05.006>
- Echaniz, E., Ho, C., Rodriguez, A., dell'Olio, L., 2019. Modelling user satisfaction in public transport systems considering missing information. *Transportation (Amst)*. <https://doi.org/10.1007/s11116-019-09996-4>
- Efthymiou, D., Antoniou, C., Tyrinopoulos, Y., Skaltsogianni, E., 2018. Factors affecting bus users' satisfaction in times of economic crisis. *Transp. Res. Part A Policy Pract.* 114, 3–12. <https://doi.org/10.1016/j.tra.2017.10.002>
- Eiró, T., Martínez, L.M., 2014. Modelling Daily Mobility Satisfaction Using a Structural Equation Model. Springer International Publishing, pp. 391–403. https://doi.org/10.1007/978-3-319-04630-3_29
- European Commission, E., 2011. Roadmap to a Single European Transport Area: Towards a Competitive and Resource Efficient Transport System: White Paper. Publications Office of the European Union, Luxembourg.
- Fellessen, M., Friman, M., 2009. Service Supply and Customer Satisfaction in Public Transportation: The Quality Paradox. *J. Public Transp.* 12, 13.
- Fellessen, M., Friman, M., 2008. Perceived satisfaction with public transport service in Nine European cities. *Transp. Res. Forum* 47, 874770. <https://doi.org/10.5399/osu/jtrf.47.3.2126>
- Filipović, S., Tica, S., Živanović, P., & Milovanović, B., Filipović, S., Tica, S., Živanović, P., Milovanović, B., 2009. Comparative analysis of the basic

- features of the expected and perceived quality of mass passenger public transport service in Belgrade. *Transport* 24, 265–273. <https://doi.org/10.3846/1648-4142.2009.24.265-273>
- Finn, A., Louviere, J.J., 1992. Determining the Appropriate Response to Evidence of Public Concern: The Case of Food Safety. *J. Public Policy Mark.* 11, 12–25. <https://doi.org/10.1177/074391569201100202>
- Friman, M., 2004. Implementing Quality Improvements in Public Transport. *J. Public Transp.* 7, 49–65. <https://doi.org/DOI:10.5038/2375-0901.7.4.3>
- Garrido, C., de Oña, R., de Oña, J., 2014. Neural networks for analyzing service quality in public transportation. *Expert Syst. Appl.* 41, 6830–6838. <https://doi.org/10.1016/j.eswa.2014.04.045>
- Garrido, R.A., Ortúzar, J.D.D., 1994. Deriving Public Transport Level of Service Weights from a Multiple Comparison of Latent and Observable Variables. *J. Oper. Res. Soc.* 45, 1099–1107. <https://doi.org/10.1057/jors.1994.180>
- Gilks, W.R. (Wally R.), Richardson, S. (Sylvia), Spiegelhalter, D.J., 1996. Markov chain Monte Carlo in practice. Chapman & Hall.
- Givoni, M., Rietveld, P., 2007. The access journey to the railway station and its role in passengers' satisfaction with rail travel. *Transp. Policy* 14, 357–365. <https://doi.org/10.1016/j.tranpol.2007.04.004>
- Gonzalo-Orden, H., dell'Olio, L., Ibeas, A., Rojo, M., 2011. Modelling gender perception of quality in interurban bus services. *Proc. ICE - Transp.* 164, 43–53. <https://doi.org/10.1680/tran.9.00031>
- Graham, J.W., Cumsille, P.E., Elek-Fisk, E., 2003. Methods for Handling Missing Data, in: *Handbook of Psychology*. John Wiley & Sons, Inc., Hoboken, NJ, USA, pp. 87–114. <https://doi.org/10.1002/0471264385.wei0204>
- Graham, J.W., Hofer, S.M., MacKinnon, D.P., 1996. Maximizing the Usefulness of Data Obtained with Planned Missing Value Patterns: An Application of Maximum Likelihood Procedures. *Multivariate Behav. Res.* 31, 197–218. https://doi.org/10.1207/s15327906mbr3102_3
- Graham, J.W., Olchowski, A.E., Gilreath, T.D., 2007. How Many Imputations are Really Needed? Some Practical Clarifications of Multiple Imputation Theory. *Prev. Sci.* 8, 206–213. <https://doi.org/10.1007/s11121-007-0070-9>
- Graham, J.W., Schafer, J.L., 1999. On the performance of multiple imputation for multivariate data with small sample size. *Stat. Strateg. small sample Res.* 50, 1–27.
- Greene, W., 2007. Limdep computer program: Version 9. Plainview, NY Econom. Softw.
- Greene, W.H., 2016. Nlogit 6 guide.
- Greene, W.H., 2008. *Econometric Analysis*, 6th edn. Prentice Hall.
- Greene, William H, Hensher, D. a., 2010. Modeling Ordered Choices: A Primer, *Modeling Ordered Choices: A Primer*. <https://doi.org/10.1017/CBO9780511845062>
- Greene, William H., Hensher, D.A., 2010. Ordered choices and heterogeneity in attribute processing. *J. Transp. Econ. Policy* 44, 331–364.

- Grujičić, D., Ivanović, I., Jović, J., Đorić, V., Dorić, V., 2014. Customer perception of service quality in public transport. *Transport* 29, 285–295. <https://doi.org/10.3846/16484142.2014.951685>
- Guirao, B., Eugenia López, M., Comendador, J., 2015. New QR Survey Methodologies to Analyze User Perception of Service Quality in Public Transport: The Experience of Madrid. *J. Public Transp.* 18.
- Guirao, B., García-Pastor, A., López-Lambas, M.E.M.E., García-Pastor, A., López-Lambas, M.E., García-Pastor, A., López-Lambas, M.E.M.E., 2016. The importance of service quality attributes in public transportation: Narrowing the gap between scientific research and practitioners' needs. *Transp. Policy* 49, 68–77. <https://doi.org/10.1016/j.tranpol.2016.04.003>
- Hapsari, R., Clemes, M., Dean, D., 2016. The Mediating Role of Perceived Value on the Relationship between Service Quality and Customer Satisfaction: Evidence from Indonesian Airline Passengers. *Procedia Econ. Financ.* 35, 388–395. [https://doi.org/10.1016/S2212-5671\(16\)00048-4](https://doi.org/10.1016/S2212-5671(16)00048-4)
- Henrickson, K., Zou, Y., Wang, Y., 2015. Flexible and Robust Method for Missing Loop Detector Data Imputation. *Transp. Res. Rec. J. Transp. Res. Board* 2527, 29–36. <https://doi.org/10.3141/2527-04>
- Hensher, D. a, 2003. Contract areas and service quality issues in public transit provision: some thoughts on the European and Australian context. *J. Public Transp.* 6, 15–42. [https://doi.org/10.1016/S0739-8859\(06\)18005-1](https://doi.org/10.1016/S0739-8859(06)18005-1)
- Hensher, D.A., 2006. Revealing differences in willingness to pay due to the dimensionality of stated choice designs: an initial assessment. *Environ. Resour. Econ.* 34, 7–44. <https://doi.org/10.1007/s10640-005-3782-y>
- Hensher, D.A., Mulley, C., 2014. Modal image: candidate drivers of preference differences for BRT and LRT. *Transportation (Amst.)* 42, 7–23. <https://doi.org/10.1007/s11116-014-9516-7>
- Hensher, D.A., Mulley, C., Yahya, N., 2010. Passenger experience with quality-enhanced bus service: The tyne and wear “superoute” services. *Transportation (Amst.)* 37, 239–256. <https://doi.org/10.1007/s11116-009-9240-x>
- Hensher, D.A., Prioni, P., 2002. A Service Quality Index for Area-wide Contract Performance Assessment. *J. Transp. Econ. Policy* 36, 93–113.
- Hensher, D.A., Stopher, P., Bullock, P., 2003. Service quality - developing a service quality index in the provision of commercial bus contracts. *Transp. Res. Part A Policy Pract.* 37, 499–517. [https://doi.org/10.1016/S0965-8564\(02\)00075-7](https://doi.org/10.1016/S0965-8564(02)00075-7)
- Hernandez, S., Monzon, A., de Oña, R., 2016. Urban transport interchanges: A methodology for evaluating perceived quality. *Transp. Res. Part A Policy Pract.* 84, 31–43. <https://doi.org/10.1016/j.tra.2015.08.008>
- Hill, T., Westbrook, R., 1997. SWOT analysis: it's time for a product recall. *Long Range Plann.* 30, 46–52. [https://doi.org/10.1016/S0024-6301\(96\)00095-7](https://doi.org/10.1016/S0024-6301(96)00095-7)
- Horton, N.J., Lipsitz, S.R., Horton, N.J., Lipsitz, S.R., 2016. Multiple Imputation in Practice : Comparison of Software Packages for Regression Models with Missing Variables Statistical Computing Software Reviews Multiple Imputation in Practice :

- Comparison of Software Packages for Regression Models With Missing Vari 55, 244–254.
- Hrncir, J., Zilecky, P., Song, Q., Jakob, M., 2017. Practical Multicriteria Urban Bicycle Routing. *IEEE Trans. Intell. Transp. Syst.* 18, 493–504. <https://doi.org/10.1109/TITS.2016.2577047>
- Huber, J., Zwerina, K., 1996. The Importance of Utility Balance in Efficient Choice Designs. *J. Mark. Res.* 33, 307. <https://doi.org/10.2307/3152127>
- Ibeas, A., dell'Olio, L., Bordagaray, M., Ortúzar, J. de D., 2014. Modelling parking choices considering user heterogeneity. *Transp. Res. Part A Policy Pract.* 70, 41–49. <https://doi.org/10.1016/j.tra.2014.10.001>
- Ibeas, A., Dell'Olio, L., Montequín, R.B., dell'Olio, L., Montequín, R.B., 2011. Citizen involvement in promoting sustainable mobility. *J. Transp. Geogr.* 19, 475–487. <https://doi.org/10.1016/j.jtrangeo.2010.01.005>
- Joewono, T.B., Kubota, H., 2007. User satisfaction with paratransit in competition with motorization in indonesia: Anticipation of future implications. *Transportation (Amst)*. 34, 337–354. <https://doi.org/10.1007/s11116-007-9119-7>
- Kim, S.H., Chung, J.-H., 2016. Reinterpretation of the Likert Scale for Public Transportation User Satisfaction. *Transp. Res. Rec. J. Transp. Res. Board* 2541, 90–99. <https://doi.org/10.3141/2541-11>
- Kleinová, E., 2016. Does liberalization of the railway industry lead to higher technical effectiveness? *J. Rail Transp. Plan. Manag.* 6, 67–76. <https://doi.org/https://doi.org/10.1016/j.jrtpm.2016.04.002>
- König, H.-H., Bleibler, F., Friederich, H.-C., Herpertz, S., Lam, T., Mayr, A., Schmidt, F., Svaldi, J., Zipfel, S., Brettschneider, C., Hilbert, A., de Zwaan, M., Egger, N., 2018. Economic evaluation of cognitive behavioral therapy and Internet-based guided self-help for binge-eating disorder. *Int. J. Eat. Disord.* 51, 155–164. <https://doi.org/10.1002/eat.22822>
- Krueger, R.A., Casey, M.A., 2000. Focus groups: A practical guide for applied research, 3rd Editio. ed. Sage publications.
- Lai, W.T., Chen, C.F., 2011. Behavioral intentions of public transit passengers-The roles of service quality, perceived value, satisfaction and involvement. *Transp. Policy* 18, 318–325. <https://doi.org/10.1016/j.tranpol.2010.09.003>
- Li, L., Su, X., Zhang, Y., Lin, Y., Li, Z., 2015. Trend Modeling for Traffic Time Series Analysis: An Integrated Study. *IEEE Trans. Intell. Transp. Syst.* 16, 3430–3439. <https://doi.org/10.1109/TITS.2015.2457240>
- Likert, R., Roslow, S., Murphy, G., 1934. A Simple and Reliable Method of Scoring the Thurstone Attitude Scales. *J. Soc. Psychol.* <https://doi.org/10.1080/00224545.1934.9919450>
- Lipovetsky, S., Conklin, M., 2014. Best-Worst Scaling in analytical closed-form solution. *J. Choice Model.* 10, 60–68. <https://doi.org/10.1016/j.jocm.2014.02.001>
- Liu, C., Erdogan, S., Ma, T., Ducca, F.W., 2016. How to Increase Rail Ridership in Maryland: Direct Ridership Models for Policy Guidance. *J. Urban Plan. Dev.* 142,

04016017. [https://doi.org/10.1061/\(ASCE\)UP.1943-5444.0000340](https://doi.org/10.1061/(ASCE)UP.1943-5444.0000340)
- Louviere, J., Lings, I., Islam, T., Gudergan, S., Flynn, T., 2013. An introduction to the application of (case 1) best-worst scaling in marketing research. *Int. J. Res. Mark.* 30, 292–303. <https://doi.org/10.1016/j.ijresmar.2012.10.002>
- Louviere, J.J., Flynn, T.N., Marley, A.A.J., 2015. Best-worst scaling: Theory, methods and applications, *Best-Worst Scaling: Theory, Methods and Applications*. <https://doi.org/10.1017/CBO9781107337855>
- Love, H.A., Durtschi, J.A., Ruhlmann, L.M., Nelson Goff, B.S., 2018. Army Soldiers and Suicidal Thoughts: The Impact of Negative Relationship Dynamics Moderated by the Dissolution of Romantic Relationships. *J. Marital Fam. Ther.* 44, 265–276. <https://doi.org/10.1111/jmft.12252>
- Machado-León, J.L., de Oña, R., Baouni, T., de Oña, J., 2017. Railway transit services in Algiers: priority improvement actions based on users perceptions. *Transp. Policy* 53, 175–185. <https://doi.org/10.1016/j.tranpol.2016.10.004>
- Machado, J.L., de Oña, R., Diez-Mesa, F., de Oña, J., 2018. Finding service quality improvement opportunities across different typologies of public transit customers. *Transp. A Transp. Sci.* 14, 761–783. <https://doi.org/10.1080/23249935.2018.1434257>
- Marley, A.A.J., Islam, T., Hawkins, G.E., 2016. A formal and empirical comparison of two score measures for best–worst scaling. *J. Choice Model.* 21, 15–24. <https://doi.org/10.1016/J.JOCM.2016.03.002>
- Marley, A.A.J., Louviere, J.J., 2005. Some probabilistic models of best, worst, and best-worst choices. *J. Math. Psychol.* 49, 464–480. <https://doi.org/10.1016/j.jmp.2005.05.003>
- Marley, A.A.J., Pihlens, D., 2012. Models of best–worst choice and ranking among multiattribute options (profiles). *J. Math. Psychol.* 56, 24–34. <https://doi.org/10.1016/J.JMP.2011.09.001>
- Martilla, J.A., James, J.C., 1977. Importance-Performance Analysis. *J. Mark.* 41, 77. <https://doi.org/10.2307/1250495>
- McKelvey, R.D., Zavoina, W., 1975. A statistical model for the analysis of ordinal level dependent variables. *J. Math. Sociol.* 4, 103–120. <https://doi.org/10.1080/0022250X.1975.9989847>
- McKelvey, R.D., Zavoina, W., 1971. Ibm Fortran-Iv Program To Perform N-Chotomous Multivariate Probit Analysis.
- Metri, B.A., 2006. Total Quality Transportation Through Deming’s 14 Points. *J. Public Transp.* 9, 35–46. <https://doi.org/10.5038/2375-0901.9.4.3>
- Minhans, A., Shahid, S., Hassan, S.A., 2015. Assessment of bus service-quality using passengers’ perceptions. *J. Teknol.* 73. <https://doi.org/10.11113/jt.v73.4290>
- Minhans, Anil, Shahid, S., Hassan, S.A., 2015. Assessment of bus service-quality using passengers??? perceptions. *J. Teknol.* 73, 61–67. <https://doi.org/10.11113/jt.v73.4290>

- Mokonyama, M., Venter, C., 2013. Incorporation of customer satisfaction in public transport contracts - A preliminary analysis. *Res. Transp. Econ.* 39, 58–66. <https://doi.org/10.1016/j.retrec.2012.05.024>
- Mouwen, A., 2015. Drivers of customer satisfaction with public transport services. *Transp. Res. Part A Policy Pract.* 78, 1–20. <https://doi.org/10.1016/j.tra.2015.05.005>
- Mulley, C., Hensher, D.A., Rose, J., 2014. Do preferences for BRT and LRT vary across geographical jurisdictions? A comparative assessment of six Australian capital cities. *Case Stud. Transp. Policy* 2, 1–9. <https://doi.org/10.1016/j.cstp.2013.11.001>
- Nandan, S., 2010. Determinants of customer satisfaction on service quality: A study of railway platforms in India. *J. public Transp.* 13, 97–113. <https://doi.org/10.5038/2375-0901.13.1.6>
- Nathanail, E., 2008. Measuring the quality of service for passengers on the hellenic railways. *Transp. Res. Part A Policy Pract.* 42, 48–66. <https://doi.org/10.1016/j.tra.2007.06.006>
- Newgard, C.D., Malveau, S., Zive, D., Lupton, J., Lin, A., 2018. Building A Longitudinal Cohort From 9-1-1 to 1-Year Using Existing Data Sources, Probabilistic Linkage, and Multiple Imputation: A Validation Study. *Acad. Emerg. Med.* acem.13512. <https://doi.org/10.1111/acem.13512>
- Nilsson, L., Johnson, M.D., Gustafsson, A., 2001. The impact of quality practices on customer satisfaction and business results: product versus service organizations. *J. Qual. Manag.* 6, 5–27. [https://doi.org/10.1016/S1084-8568\(01\)00026-8](https://doi.org/10.1016/S1084-8568(01)00026-8)
- Parasuraman, A., Zeithaml, V.A., Berry, L.L., 1988. Servqual: A Multiple-Item Scale For Measuring Consumer Perc. *J. Retail.* 64, 12.
- Parasuraman, A., Zeithaml, V.A., Berry, L.L., Parasuraman, 1985. A conceptual model of service quality and its implications for future research. *J. Mark.* <https://doi.org/10.2307/1251430>
- Park, Y., Ha, H.-K.K., 2006. Analysis of the impact of high-speed railroad service on air transport demand. *Transp. Res. Part E Logist. Transp. Rev.* 42, 95–104. <https://doi.org/https://doi.org/10.1016/j.tre.2005.09.003>
- Pettersson, M.E., Koppelman, G.H., Flokstra-de Blok, B.M.J., Kollen, B.J., Dubois, A.E.J., 2018. Prediction of the severity of allergic reactions to foods. *Allergy* 73, 1532–1540. <https://doi.org/10.1111/all.13423>
- Phan, M.H., Keebler, J.R., Chaparro, B.S., 2016. The Development and Validation of the Game User Experience Satisfaction Scale (GUESS). *Hum. Factors J. Hum. Factors Ergon. Soc.* 58, 1217–1247. <https://doi.org/10.1177/0018720816669646>
- Pratt, J.W., 1981. Concavity of the Log Likelihood. *J. Am. Stat. Assoc.* 76, 103–106. <https://doi.org/10.1080/01621459.1981.10477613>
- Raghunathan, T.E., Lepkowski, J.M., Van Hoewyk, J., Solenberger, P., 2001. A multivariate technique for multiply imputing missing values using a sequence of regression models. *Surv. Methodol.* 27, 85–95.
- Rahman, F., Das, T., Hadiuzzaman, M., Hossain, S., 2016. Perceived service quality of paratransit in developing countries: A structural equation approach. *Transp. Res.*

- Part A Policy Pract. 93, 23–38. <https://doi.org/10.1016/j.tra.2016.08.008>
- Rea, L.M., Parker, R.A., 2014. Designing and conducting survey research: A comprehensive guide, Jossey Bass Public Administration Series. Jossey-Bass.
- Rissel, C., Crane, M., Wen, L.M., Greaves, S., Standen, C., 2016. Satisfaction with transport and enjoyment of the commute by commuting mode in inner Sydney. *Heal. Promot. J. Aust.* 27, 80–83. <https://doi.org/10.1071/HE15044>
- Ritchie, J.R.B., Johnston, E.E., Jones, V.J., 1980. Competition, Fares and Fences—Perspective of the Air Traveler. *J. Travel Res.* 18, 17–25. <https://doi.org/10.1177/004728758001800304>
- Rojó, M., dell’Olio, L., Gonzalo-Orden, H., Ibeas, Á., 2013. Interurban bus service quality from the users’ viewpoint. *Transp. Plan. Technol.* 36, 599–616. <https://doi.org/10.1080/03081060.2013.845432>
- Rojó, M., Gonzalo-Orden, H., dell’Olio, L., Ibeas, ángel, 2012. Relationship between service quality and demand for inter-urban buses. *Transp. Res. Part A Policy Pract.* 46, 1716–1729. <https://doi.org/10.1016/j.tra.2012.07.006>
- Román, C., Espino, R., Martín, J.C., 2010. Analyzing Competition between the High Speed Train and Alternative Modes. The Case of the Madrid-Zaragoza-Barcelona Corridor. *J. Choice Model.* 3, 84–108. [https://doi.org/https://doi.org/10.1016/S1755-5345\(13\)70030-7](https://doi.org/https://doi.org/10.1016/S1755-5345(13)70030-7)
- Román, C., Espino, R., Martín, J.C., 2007. Competition of high-speed train with air transport: The case of Madrid–Barcelona. *J. Air Transp. Manag.* 13, 277–284. <https://doi.org/https://doi.org/10.1016/j.jairtraman.2007.04.009>
- Román, C., Martín, J.C., Espino, R., 2014. Using Stated Preferences to Analyze the Service Quality of Public Transport. *Int. J. Sustain. Transp.* 8, 28–46. <https://doi.org/10.1080/15568318.2012.758460>
- Rose, J., Collins, A., Bliemer, M., Hensher, D., 2009. Ngene 1.0 stated choice experiment design software.
- Rose, J.M., Bliemer, M.C.J., 2014. Stated choice experimental design theory : The who , the what and the why. *Handb. Choice Model.* 152–177.
- Rose, J.M., Bliemer, M.C.J., 2013. Sample size requirements for stated choice experiments. *Transportation (Amst).* 40, 1021–1041. <https://doi.org/10.1007/s11116-013-9451-z>
- Rose, J.M., Bliemer, M.C.J., 2009. Constructing Efficient Stated Choice Experimental Designs. *Transp. Rev.* 29, 587–617. <https://doi.org/10.1080/01441640902827623>
- Rose, J.M., Hensher, D.A., 2018. User satisfaction with taxi and limousine services in the Melbourne metropolitan area. *J. Transp. Geogr.* 70, 234–245. <https://doi.org/10.1016/j.jtrangeo.2018.06.017>
- Roth, P.L., 1994. MISSING DATA: A CONCEPTUAL REVIEW FOR APPLIED PSYCHOLOGISTS. *Pers. Psychol.* 47, 537–560. <https://doi.org/10.1111/j.1744-6570.1994.tb01736.x>
- Rothfuss, P., 2008. The name of the wind. Gollancz.

- Rubin, D.B., 2004. Multiple imputation for nonresponse in surveys. Wiley-Interscience.
- Rubin, D.B., 1996. Multiple Imputation after 18+ Years. *J. Am. Stat. Assoc.* 91, 473–489. <https://doi.org/10.1080/01621459.1996.10476908>
- Rubin, D.B., 1978. Multiple imputations in sample surveys - A phenomeno- logical Bayesian approach to nonresponse, in: *Proceedings of the Survey Research Methods Section of the American Statistical Association*. pp. 20–28.
- Rubin, D.B., 1977. Formalizing Subjective Notions About the Effect of Nonrespondents in Sample Surveys Formalizing Sub jective Notions About the Effect of Nonrespondents in Sample Surveys. *Source J. Am. Stat. Assoc.* 72144202, 538–543.
- Rubin, D.B., 1976. Inference and missing data. *Biometrika* 63, 581–592. <https://doi.org/10.1093/biomet/63.3.581>
- Sanjuan, V., 2014. Lo que te hace grande : 50 cosas que aprendí corriendo por el mundo. Alienta.
- Sanko, N., 2001. Guidelines for stated preference experiment design. MBA Dissertation. Ecole Nationale des Ponts et Chaussées, Paris.
- Schafer, J.L., Graham, J.W., 2002. Missing data: Our view of the state of the art. *Psychol. Methods* 7, 147–177. <https://doi.org/10.1037/1082-989X.7.2.147>
- Scott, L.J., Freese, J., 2006. Regression models for categorical dependent variables using Stata. Stata press.
- Sever, I., 2015. Importance-performance analysis: A valid management tool? *Tour. Manag.* 48, 43–53. <https://doi.org/10.1016/J.TOURMAN.2014.10.022>
- Shen, W., Xiao, W., Wang, X., 2016. Passenger satisfaction evaluation model for Urban rail transit: A structural equation modeling based on partial least squares. *Transp. Policy* 46, 20–31. <https://doi.org/10.1016/j.tranpol.2015.10.006>
- St-Louis, E., Manaugh, K., Van Lierop, D., El-Geneidy, A., 2014. The happy commuter: A comparison of commuter satisfaction across modes. *Transp. Res. Part F Traffic Psychol. Behav.* 26, 160–170. <https://doi.org/10.1016/j.trf.2014.07.004>
- Sterne, J.A.C., White, I.R., Carlin, J.B., Spratt, M., Royston, P., Kenward, M.G., Wood, A.M., Carpenter, J.R., 2009. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls. *BMJ* 338, b2393. <https://doi.org/10.1136/bmj.b2393>
- Sun, X., Zhang, Y., Wandelt, S., 2017. Air Transport versus High-Speed Rail: An Overview and Research Agenda. *J. Adv. Transp.* 2017, 1–18. <https://doi.org/10.1155/2017/8426926>
- Susilo, Y.O., Cats, O., 2014. Exploring key determinants of travel satisfaction for multi-modal trips by different traveler groups. *Transp. Res. Part A Policy Pract.* 67, 366–380. <https://doi.org/10.1016/j.tra.2014.08.002>
- Tam, M.L., Lam, W.H.K., Lo, H.P., 2010. Incorporating passenger perceived service quality in airport ground access mode choice model. *Transportmetrica* 6, 3–17. <https://doi.org/10.1080/18128600902929583>
- Tang, J., Wang, Y., Zhang, S., Wang, H., Liu, F., Yu, S., 2015. On Missing Traffic Data

- Imputation Based on Fuzzy *C* -Means Method by Considering Spatial–Temporal Correlation. *Transp. Res. Rec. J. Transp. Res. Board* 2528, 86–95. <https://doi.org/10.3141/2528-10>
- Troyanskaya, O., Cantor, M., Sherlock, G., Brown, P., Hastie, T., Tibshirani, R., Botstein, D., Altman, R.B., 2001. Missing value estimation methods for DNA microarrays. *Bioinformatics* 17, 520–525. <https://doi.org/10.1093/bioinformatics/17.6.520>
- Tsami, M., Nathanail, E., 2017. Guidance Provision for Increasing Quality of Service of Public Transport, in: *Procedia Engineering*. pp. 551–557. <https://doi.org/10.1016/j.proeng.2017.01.108>
- Tyrinopoulos, Y., Antoniou, C., 2015. Analysis of passengers' perception of public transport quality and performance, *Transportation Systems and Engineering: Concepts, Methodologies, Tools, and Applications*. <https://doi.org/10.4018/978-1-4666-8473-7.ch056>
- Tyrinopoulos, Y., Antoniou, C., 2008. Public transit user satisfaction: Variability and policy implications. *Transp. Policy* 15, 260–272. <https://doi.org/10.1016/j.tranpol.2008.06.002>
- van Buuren, S., 2007. Multiple imputation of discrete and continuous data by fully conditional specification. *Stat. Methods Med. Res.* 16, 219–242. <https://doi.org/10.1177/0962280206074463>
- van Buuren, S., Boshuizen, H.C., Knook, D.L., 1999. Multiple imputation of missing blood pressure covariates in survival analysis. *Stat. Med.* 18, 681–694. [https://doi.org/10.1002/\(SICI\)1097-0258\(19990330\)18:6<681::AID-SIM71>3.0.CO;2-R](https://doi.org/10.1002/(SICI)1097-0258(19990330)18:6<681::AID-SIM71>3.0.CO;2-R) [pii]
- Van Buuren, S., Brand, J.P.L., Groothuis-Oudshoorn, C.G.M., Rubin, D.B., 2006. Fully conditional specification in multivariate imputation 76, 1049–1064. <https://doi.org/10.1080/10629360600810434>
- Vicente, P., Reis, E., 2016. Profiling public transport users through perceptions about public transport providers and satisfaction with the public transport service. *Public Transp.* 8, 387–403. <https://doi.org/10.1007/s12469-016-0141-z>
- Vuong, Q.H., 1989. Likelihood Ratio Tests for Model Selection and Non-Nested Hypotheses. *Econometrica* 57, 307. <https://doi.org/10.2307/1912557>
- Ward, H., Atkins, J., 2002. From their lives: A manual on how to conduct focus groups of low-income parents. Institute for Child and Family Policy, University of Southern Maine, Portland, Maine.
- Weinstein, A., 2000. Customer Satisfaction Among Transit Riders: How Customers Rank the Relative Importance of Various Service Attributes. *Transp. Res. Rec. J. Transp. Res. Board* 1735, 123–132. <https://doi.org/10.3141/1735-15>
- Wellington, C., Aylwin, M., 2013. A life without limits: A World Champion's Journey. Little, Brown Book Group.
- Wen, C.-H., Lan, L.W., Chen, C.-H., 2005. Passengers Perception on Service Quality and Their Choice for Intercity Bus Services. *Transp. Res. Board*, 84th Annu. Meet.
- Wongwiriya, P., Nakamura, F., Tanaka, S., Ariyoshi, R., 2017. User Satisfaction of

- Songtaew in Thailand: Case Study of Khon Kaen City, in: Transportation Research Procedia. Elsevier B.V., pp. 4946–4953. <https://doi.org/10.1016/j.trpro.2017.05.372>
- Zou, M., 2015. Gender, work orientations and job satisfaction. Work. Employ. Soc. 29, 3–22. <https://doi.org/10.1177/0950017014559267>

